

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
27 May 2004 (27.05.2004)

PCT

(10) International Publication Number
WO 2004/044178 A2

(51) International Patent Classification⁷: C12N

(21) International Application Number:
PCT/US2003/036260

(22) International Filing Date:
13 November 2003 (13.11.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/425,813 13 November 2002 (13.11.2002) US

(71) Applicant (*for all designated States except US*): **GENENTECH, INC.** [US/US]; 1 DNA Way, South San Francisco, CA 94080-4990 (US).

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): **SMITH, Victoria** [AU/US]; 19 Dwight Road, Burlingame, CA 94010 (US).

(74) Agents: **CONLEY, Deirdre L.** et al.; GENENTECH, INC., 1 DNA Way, South San Francisco, CA 94080-4990 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND COMPOSITIONS FOR DIAGNOSING DYSPLASIA

(57) Abstract: Methods and compositions are disclosed for detecting dysplasia in a tissue sample, screening candidate compounds for the ability to inhibit growth of a cancer cell, predicting predisposition to adenocarcinoma and treating cancer based on gene expression profiles.



WO 2004/044178 A2

5

METHODS AND COMPOSITIONS FOR DETECTING DYSPLASIA

TECHNICAL FIELD

10

The present invention relates to nucleic acid sequences, and compositions and uses therefore, which have been shown to be differentially expressed in high-grade dysplasia and which are useful as markers for the detection of high-grade dysplasia in a patient, and are implicated in the development of adenocarcinoma.

15

BACKGROUND OF THE INVENTION

The incidence of esophageal adenocarcinoma is rising in Western Countries, replacing squamous cell carcinoma as the most common neoplasm of the esophagus in white males and increasing in other ethnic groups (Devesa et al., Cancer 83:2049-2053 (1998); and
20 Bollschweiler et al., Cancer 92:549-555 (2001)). Barrett's esophagus (BE) is the primary recognized risk factor for esophageal adenocarcinoma. BE results from repeated injury to the esophageal mucosa and develops in a subset of patients with chronic gastrointestinal reflux disease. It is characterized by a metaplastic change of squamous esophageal epithelium to intestinalized columnar mucosa (Csendes et al., Dis. Esoph 13:5-11 (2000); Cameron et al.,
25 New Eng. J. Med. 313:857-859 (1985); and Drewitz et al., Amer. J. Gastroenterol 92:212-215 (1997)).

Barrett's esophagus is found in 6% -16% of patients undergoing upper gastrointestinal endoscopy for gastroesophageal reflux, and it is estimated that a substantial patient population
30 remains undiagnosed (Sarr et al., Amer. J. Surgery 149:187-193 (1985); Winters et al., Gastroenterology 92:118-124 (1985); Cameron et al., Gastroenterology 99:918-922 (1990); and Cameron et al., Gastroenterology 103:1241-1245 (1992)). The risk of developing esophageal carcinoma is 30 – 150 times greater in patients with BE. The outlook for patients diagnosed with adenocarcinoma is poor, with a 5 year survival rate of 10 – 15% (Streitz et al.,

Ann. Surg. 213:122-125 (1991); Menke-Pluymers et al., Gut 33:1454-1458 (1992); and Lerut et al., J. Thorac. Cardiovasc. Surg. 107:1059-1066 (1994)). Patients with BE are placed on surveillance programs, although the absolute risk of developing adenocarcinoma in the context of BE remains relatively low, estimated at approximately 0.5% per patient year (Drewitz et al., Amer. J. Gastroenterol 92:212-215; O'Connor et al., Am. J. Gastroenterol 94:2037-2042 (1999); Spechler et al., JAMA 285:2331-2338 (2001); and Shaheen et al., Gastroenterology 119:333-338 (2000)). The value and cost-effectiveness of surveillance programs continue to be debated due to lack of understanding of the natural history of BE, the difficulty in obtaining representative biopsies by random sampling due to the heterogeneous nature of intestinal metaplasia, and inter-observer variability in endoscopic and histopathologic diagnosis (Falk, Gastroenterology 122:1569-1591 (2002); Sampliner, Am. J Gastroenterol. 93:1028-1032 (1998); and Alikhan et al., Gastrointest. Endosc. 50:23-26 (1999)). A metaplasia-dysplasia-carcinoma sequence has been described for BE and genetic changes involving cell cycle abnormalities, DNA ploidy, mutations, and amplification and expression of oncogenes have been identified (al-Kasspooles et al., Internat. J. Cancer 54:213-219 (1993); Vissers et al., Anticancer Res. 21:3813-3820 (2001); Bani-Hani et al., J. Natl. Cancer Inst. 92:1316-1321 (2000); Walch et al., Am. J. Pathol. 156:555-566 (2000); Wong et al., Cancer Res. 61:8284-8289 (2001); and Romagnoli et al., Laboratory Investigation 81:241-247 (2001)). There is a need for reliable detection of high-grade dysplasia and diagnosis of patients, such as BE patients, likely to develop adenocarcinoma, thereby allowing the disease to be monitored and treated early in its progression.

SUMMARY OF THE INVENTION

Generally, the present invention is based on the discovery that it is possible to detect high-grade dysplasia in a patient suspected of experiencing dysplasia, such as dysplasia associated with gastrointestinal reflux disease, such as Barrett's esophagus, or colon tissue dysplasia, by determining expression in an esophageal or colon biopsy from the patient wherein at least eight genes selected from a group of genes are expressed at a level of at least 1.5 fold over expression in a control sample. The control sample may comprise an esophageal or colon biopsy from a normal patient (i.e. one not experiencing gastrointestinal reflux disease) or from pooled samples of normal epithelial tissue (such as from normal liver, lung and kidney tissue). The group of high-grade dysplasia (HGD) gene markers, and their encoded polypeptides, comprise ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1 or 2);

AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3 or 4); ADAM8 (NM_001109) (SEQ ID NO:5 or 6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7 or 8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9 or 10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11 or 12); TM7SF1 (NM_003272) (SEQ ID NO:13 or 14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15 or 16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17 or 18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19 or 20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21 or 22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23 or 24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25 or 26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27 or 28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29 or 30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31 or 32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33 or 34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35 or 36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37 or 38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39 or 40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41 or 42); and TCF4 (NM_030756) (SEQ ID NO:43 or 44). HGD marker polypeptides refer to the polypeptides encoded by the HGD gene markers.

In an aspect, the invention involves a method for the diagnosis of esophageal high-grade dysplasia (HGD) in a patient, comprising establishing increased expression of at least eight genes (listed here with the polypeptide encoded by the gene) selected from the group consisting of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1 or 2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3 or 4); ADAM8 (NM_001109) (SEQ ID NO:5 or 6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7 or 8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9 or 10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11 or 12); TM7SF1 (NM_003272) (SEQ ID NO:13 or 14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15 or 16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17 or 18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19 or 20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21 or 22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23 or 24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25 or 26); PA21 (phospholipase a2 precursor,

NM_000928) (SEQ ID NO:27 or 28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29 or 30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31 or 32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33 or 34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35 or 36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37 or 38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39 or 40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41 or 42); and TCF4 (NM_030756) (SEQ ID NO:43 or 44); and comparing expression of the genes to a baseline expression of the genes in normal tissue controls; wherein an increase of at least 1.5-fold in expression (and/or p value < 0/07) of the genes from the group relative to the baseline indicates that the patient is experiencing esophageal high-grade dysplasia. In an embodiment of the invention, the tissue is human tissue.

In another embodiment, the invention involves a method of identifying a patient susceptible to esophageal adenocarcinoma, comprising diagnosing esophageal high-grade dysplasia in a patient by establishing increased expression of at least eight genes selected from the group consisting of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3); ADAM8 (NM_001109) (SEQ ID NO:5); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11); TM7SF1 (NM_003272) (SEQ ID NO:13); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41); and TCF4 (NM_030756) (SEQ ID NO:43); and comparing expression of the genes to a baseline expression of the genes in

normal tissue controls; wherein an increase of at least 1.5-fold in expression of the genes from the group relative to the baseline indicates that the patient is experiencing esophageal high-grade dysplasia. Alternatively, the patient may be susceptible to colon carcinoma and the diagnosing of high-grade dysplasia is by similarly determining expression of at least eight
 5 genes of the above group in a test colon tissue sample compared to a normal colon tissue sample.

In still another embodiment, the invention involves a method for determining whether an esophageal tissue is predisposed to a neo-plastic transformation, comprising determining
 10 whether in a cell from the esophageal tissue at least eight nucleic acid sequences selected from the group consisting of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3); ADAM8 (NM_001109) (SEQ ID NO:5); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9); NROB2 (Nuclear hormone
 15 receptor, NM_021969) (SEQ ID NO:11); TM7SF1 (NM_003272) (SEQ ID NO:13); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21); SLNAC1 (sodium channel receptor SLNAC1, NM_004769)
 20 (SEQ ID NO:23); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35); PHYH
 25 (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41); and TCF4 (NM_030756) (SEQ ID NO:43) is expressed at least 1.5-fold above baseline expression in a normal tissue control. In an embodiment, the tissue is human tissue.

In another aspect, the invention involves a method for the diagnosis of esophageal high-grade dysplasia in a patient, comprising establishing the level of expression a polypeptide encoded by at least eight genes selected from the group consisting of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog,

NM_006408) (SEQ ID NO:3); ADAM8 (NM_001109) (SEQ ID NO:5); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11); TM7SF1 (NM_003272) (SEQ ID NO:13); DLDH (dihydrolipamide dehydrogenase, 5 NM_000108) (SEQ ID NOS:15); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25); PA21 (phospholipase a2 precursor, 10 NM_000928) (SEQ ID NO:27); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37); CYB5 (cytochrome b5, 3' end, 15 NM_001914) (SEQ ID NO:39); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41); and TCF4 (NM_030756) (SEQ ID NO:43); and comparing expression of the at least eight genes from the group to a baseline expression of the genes in normal tissue controls; wherein an increase of at least 1.5-fold in expression of the polypeptide encoded by the genes from the group relative to the baseline indicates that the patient has 20 esophageal dysplasia.

In an embodiment, the method involves contacting a HGD cell or a cancer cell with an antibody that binds specifically to a polypeptide, or fragment thereof, encoded by a gene selected from the group of HGD marker genes or cancer marker genes as disclosed herein.

25

In an embodiment, the method involves determining expression of at least 8 of the genes of the group of HGD marker genes using by nucleic acid microarray analysis. In further embodiment, the microarray comprises nucleic acid sequences of at least 20 nucleotides derived from at least eight of the genes from the following group: ET-1 (endothelin-1, 30 NM_001955) (SEQ ID NO:1); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3); ADAM8 (NM_001109) (SEQ ID NO:5); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11); TM7SF1 (NM_003272) (SEQ ID NO:13); DLDH (dihydrolipamide dehydrogenase,

NM_000108) (SEQ ID NOS:15); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41); and TCF4 (NM_030756) (SEQ ID NO:43).

In another embodiment, the invention involves analysis using a microarray comprising nucleic acid probe sequences comprising at least 20 contiguous nucleotides from at least 8 genes selected from the group of HGD marker genes: ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3); ADAM8 (NM_001109) (SEQ ID NO:5); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11); TM7SF1 (NM_003272) (SEQ ID NO:13); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41); and TCF4 (NM_030756) (SEQ ID NO:43).

In a further embodiment, the methods of detecting high-grade dysplasia, diagnosing high-grade dysplasia, or prognosing development of cancer from detected high-grade dysplasia involves determining expression of at least eight genes from the group of HGD markers disclosed herein above as determined by an analysis method including, but not limited to polymerase chain reaction analysis, real-time polymerase chain reaction analysis, Taqman® polymerase chain reaction analysis, nucleic acid hybridization, fluorescent *in situ* hybridization and non-fluorescent *in situ* hybridization (e.g. radioactive, calorimetric, enzymatic or enzyme-linked detection methods for *in situ* hybridization). Where the method of the invention involves determining increased expression of polypeptides encoded by at least eight HGD marker genes as disclosed herein above, an embodiment of the method involves analysis using an antibody capable of specifically binding to a polypeptide, or a fragment thereof, encoded by a HGD marker gene.

In an alternative embodiment, the analytical methods of the invention involve probes or targets labelled with radionuclides or enzymatic labels such that expression of a gene or polypeptide is determinable.

In an embodiment of any of the methods or compositions of the invention, the dysplasia is high-grade dysplasia of esophagus tissue and the cancer is esophageal adenocarcinoma. Alternatively the patient is a human patient.

In another aspect, the invention involves a method of treating high-grade esophageal dysplasia or inhibiting or preventing cancer in a patient in need of such treatment, the method comprising administering to the patient a compound capable of decreasing expression of a gene selected from the group consisting of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3); ADAM8 (NM_001109) (SEQ ID NO:5); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11); TM7SF1 (NM_003272) (SEQ ID NO:13); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ

ID NO:25); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35);
5 PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41); and TCF4 (NM_030756) (SEQ ID NO:43) .

10 In still another aspect, the invention involves a method of treating high-grade esophageal dysplasia or inhibiting or preventing cancer in a patient in need of such treatment, the method comprising administering to the patient a compound capable of decreasing expression of a polypeptide encoded by a gene selected from the HGD marker genes.

15 In still another aspect, the invention involves a method of treating high-grade esophageal dysplasia or inhibiting or preventing cancer in a patient in need of such treatment, the method comprising administering to the patient a compound capable of inhibiting activity of a polypeptide encoded by a gene which is one of at least eight genes selected from the group of HGD marker genes as disclosed herein. In an embodiment, the compound is an
20 antagonist of the polypeptide. In a further embodiment, the antagonist is an antibody, such as a monoclonal antibody or a humanized monoclonal antibody.

In a further aspect, the invention involves a method of screening for candidate drugs which inhibits or prevents progression from dysplasia to adenocarcinoma, the method
25 comprising contacting a cell with a candidate drug, and assaying inhibition of progression from high-grade dysplasia to cancer in the cell, wherein the cell, prior to contacting with the candidate drug, expresses at least eight genes at a level at least 1.5-fold increased relative to a normal tissue baseline level, wherein the genes are selected from group of HGD marker genes as disclosed herein.

30

In another aspect, the invention involves a method of inhibiting or preventing progression from high-grade dysplasia to cancer in a patient by administering a drug identified by screening for candidate drugs which inhibits or prevents progression from dysplasia to adenocarcinoma, the method comprising contacting a cell with a candidate drug, and assaying

inhibition of progression from high-grade dysplasia to cancer in the cell, wherein the cell, prior to contacting with the candidate drug, expresses at least eight genes at a level at least 1.5-fold increased relative to a normal tissue baseline level, wherein the genes are selected from group of HGD marker genes as disclosed herein.

5

In another aspect, the invention involves a compound capable of inhibiting or preventing the progression from high-grade dysplasia to cancer in a patient. In an embodiment of the invention the compound is identified by screening for a candidate drug which inhibits or prevents progression from dysplasia to adenocarcinoma, the method comprising contacting a
10 cell expressing at least 1.5-fold relative to a normal tissue baseline level at least eight genes selected from the group of HGD marker genes as disclosed herein, with a candidate drug, and assaying inhibition of progression from high-grade dysplasia to cancer in the cell. In an embodiment, the invention involves a pharmaceutical composition comprising a compound
15 capable of inhibiting or preventing the progression from high-grade dysplasia to cancer in a patient, and a pharmaceutically acceptable carrier.

In still another aspect, the invention involves detecting cancer in a patient by determining that a gene, or the polypeptide it encodes, selected from the group consisting of CAD17 (liver-intestine cadherin, NM_004063) (SEQ ID NO:45 or 46), CLDN15 (claudin 15,
20 NM_014343) (SEQ ID NO:47 or 48), SLNAC1 (sodium channel, NM_004769) (SEQ ID NO:23 or 24), CFTR (chloride channel, NM_000492) (SEQ ID NO:49 or 50), H2R (histamine H2 receptor, NM_022304) (SEQ ID NO:51 or 52), PRSS8 (serine protease, NM_002773) (SEQ ID NO:7 or 8), PA21 (phospholipase A2 group IB, NM_000928) (SEQ ID NO:27 or 28), AGR2 (anterior gradient 2 homolog, (NM_006408) (SEQ ID NO:3 or 4), EGFR
25 (NM_005228) (SEQ ID NO:53 or 54), EPHB2 (NM_004442) (SEQ ID NO:55 or 56), CRIPTO CR-1 (NM_003212) (SEQ ID NO:57 or 58), Eprin B1 (NM_004429) (SEQ ID NO:59 or 60), MMP-17/MT4-MMP (NM_016155) (SEQ ID NO:61 or 62), MMP26 (NM_021801) (SEQ ID NO:63 or 64), ADAM10 (NM_001110) (SEQ ID NO:65 or 66), ADAM8 (NM_001109) (SEQ ID NO:5 or 6), ADAM1 (XM_132370) (SEQ ID NO:67 or 68),
30 TIM1 (NM_003254) (SEQ ID NO:69 or 70), MUC1 (XM_053256) (SEQ ID NO:71 or 72), CEA (NM_004363) (SEQ ID NO:73 or 74), NCA (NM_002483) (SEQ ID NO:75 or 76), Follistatin (NM_006350) (SEQ ID NO:77 or 78), Claudin 1 (NM_021101) (SEQ ID NO:79 or 80), Claudin 14 (NM_012130) (SEQ ID NO:81 or 82), tenascin-R (NM_003285) (SEQ ID NO:83 or 84), CAD3 (NM_001793) (SEQ ID NO:85 or 86), AXO1 (NM_005076) (SEQ ID

NO:9 or 10), CONT (NM_001843) (SEQ ID NO:87 or 88), Osteopontin (NM_000582) (SEQ ID NO:89 or 90), Galectin 8 (NM_006499) (SEQ ID NO:91 or 92), PGS1 (bilycan, NM_001711) (SEQ ID NO:93 or 94), Frizzled 2 (NM_001466) (SEQ ID NO:95 or 96), ISLR (NM_005545) (SEQ ID NO:97 or 98), FLJ23399 (NM_022763) (SEQ ID NO:99 or 100),
 5 TEM1 (NM_020404) (SEQ ID NO:101 or 102), Tie2 ligand2 (NM_001147) (SEQ ID NO:103 or 104), STC-2 (NM_003714) (SEQ ID NO:19 or 20), VEGFC (NM_005429) (SEQ ID NO:105 or 106), tPA (NM_000930) (SEQ ID NO:107 or 108), Endothelin 1 (NM_001955) (SEQ ID NO:1 or 2), Thrombomodulin (NM_000361) (SEQ ID NO:109 or 110), TF (NM_001993) (SEQ ID NO:111 or 112), GPR4 (NM_005282) (SEQ ID NO:113 or 114),
 10 GPR66 (NM_006056) (SEQ ID NO:115 or 116), SLC22A2 (NM_003058) ((SEQ ID NO:117 or 118), MLSN1 (NM_002420) (SEQ ID NO:119 or 120), and ATN2 (Na/K transport, NM_000702) (SEQ ID NO:121 or 122) is expressed at a level of about 1.5-fold in a test sample above the level of expression in a normal tissue sample of the same tissue type. The test sample is generally from a patient suspected of experiencing cancer, including, but not
 15 limited to, adenocarcinoma, esophageal adenocarcinoma, or colon cancer. The test sample is generally from the esophagus or colon of the patient. In an embodiment, at least two, alternatively at least three, alternatively at least five, and alternatively at least eight genes selected from the above group is upregulated in cancer tissue at 1.5-fold relative to normal tissue. Detection of the up-regulation of these genes is determined by, for example,
 20 hybridization analysis as standard in the and disclosed herein, as well as through antibody binding analysis of the level polypeptides expressed by the up-regulated gene or genes.

In an embodiment, the invention involves treatment by contacting a cancer cell with a compound that inhibits expression of at least one, optionally at least two, at least three, at least
 25 five, or at least eight genes, or the polypeptides encoded by the genes, selected from the group consisting of CAD17 (liver-intestine cadherin, NM_004063) (SEQ ID NO:45 or 46), CLDN15 (claudin 15, NM_014343) (SEQ ID NO:47 or 48), SLNAC1 (sodium channel, NM_004769) (SEQ ID NO:23 or 24), CFTR (chloride channel, NM_000492) (SEQ ID NO:49 or 50), H2R (histamine H2 receptor, NM_022304) (SEQ ID NO:51 or 52), PRSS8 (serine protease,
 30 NM_002773) (SEQ ID NO:7 or 8), PA21 (phospholipase A2 group IB, NM_000928) (SEQ ID NO:27 or 28), AGR2 (anterior gradient 2 homolog, (NM_006408) (SEQ ID NO:3 or 4), EGFR (NM_005228) (SEQ ID NO:53 or 54), EPHB2 (NM_004442) (SEQ ID NO:55 or 56), CRIPTO CR-1 (NM_003212) (SEQ ID NO:57 or 58), Eprin B1 (NM_004429) (SEQ ID NO:59 or 60), MMP-17/MT4-MMP (NM_016155) (SEQ ID NO:61 or 62), MMP26

(NM_021801) (SEQ ID NO:63 or 64), ADAM10 (NM_001110) (SEQ ID NO:65 or 66), ADAM8 (NM_001109) (SEQ ID NO:5 or 6), ADAM1 (XM_132370) (SEQ ID NO:67 or 68), TIM1 (NM_003254) (SEQ ID NO:69 or 70), MUC1 (XM_053256) (SEQ ID NO:71 or 72), CEA (NM_004363) (SEQ ID NO:73 or 74), NCA (NM_002483) (SEQ ID NO:75 or 76),
 5 Follistatin (NM_006350) (SEQ ID NO:77 or 78), Claudin 1 (NM_021101) (SEQ ID NO:79 or 80), Claudin 14 (NM_012130) (SEQ ID NO:81 or 82), tenascin-R (NM_003285) (SEQ ID NO:83 or 84), CAD3 (NM_001793) (SEQ ID NO:85 or 86), AXO1 (NM_005076) (SEQ ID NO:9 or 10), CONT (NM_001843) (SEQ ID NO:87 or 88), Osteopontin (NM_000582) (SEQ ID NO:89 or 90), Galectin 8 (NM_006499) (SEQ ID NO:91 or 92), PGS1 (bilycan,
 10 NM_001711) (SEQ ID NO:93 or 94), Frizzled 2 (NM_001466) (SEQ ID NO:95 or 96), ISLR (NM_005545) (SEQ ID NO:97 or 98), FLJ23399 (NM_022763) (SEQ ID NO:99 or 100), TEM1 (NM_020404) (SEQ ID NO:101 or 102), Tie2 ligand2 (NM_001147) (SEQ ID NO:103 or 104), STC-2 (NM_003714) (SEQ ID NO:19 or 20), VEGFC (NM_005429) (SEQ ID NO:105 or 106), tPA (NM_000930) (SEQ ID NO:107 or 108), Endothelin 1 (NM_001955)
 15 (SEQ ID NO:1 or 2), Thrombomodulin (NM_000361) (SEQ ID NO:109 or 110), TF (NM_001993) (SEQ ID NO:111 or 112), GPR4 (NM_005282) (SEQ ID NO:113 or 114), GPR66 (NM_006056) (SEQ ID NO:115 or 116), SLC22A2 (NM_003058) ((SEQ ID NO:117 or 118), MLSN1 (NM_002420) (SEQ ID NO:119 or 120), and ATN2 (Na/K transport, NM_000702) (SEQ ID NO:121 or 122). In another embodiment, treatment is by contacting
 20 the cancer cell with a compound that inhibits the production or activity of a polypeptide of the above group and/or encoded by a gene of the above group. Where inhibition of a polypeptide is desired, the compound is often an antibody specific for the polypeptide, is often a monoclonal antibody such as a humanized antibody.

25 In yet another aspect, the invention involves a method of screening a candidate compound for the ability to inhibit cancer cell growth or cause cancer cell death by contacting the candidate compound with a cancer cell expressing a gene or polypeptide selected from the following group: CAD17 (liver-intestine cadherin, NM_004063) (SEQ ID NO:45 or 46), CLDN15 (claudin 15, NM_014343) (SEQ ID NO:47 or 48), SLNAC1 (sodium channel,
 30 NM_004769) (SEQ ID NO:23 or 24), CFTR (chloride channel, NM_000492) (SEQ ID NO:49 or 50), H2R (histamine H2 receptor, NM_022304) (SEQ ID NO:51 or 52), PRSS8 (serine protease, NM_002773) (SEQ ID NO:7 or 8), PA21 (phospholipase A2 group IB, NM_000928) (SEQ ID NO:27 or 28), AGR2 (anterior gradient 2 homolog, NM_006408) (SEQ ID NO:3 or 4), EGFR (NM_005228) (SEQ ID NO:53 or 54), EPHB2 (NM_004442) (SEQ ID NO:55 or

56), CRIPTO CR-1 (NM_003212) (SEQ ID NO:57 or 58), Eprin B1 (NM_004429) (SEQ ID NO:59 or 60), MMP-17/MT4-MMP (NM_016155) (SEQ ID NO:61 or 62), MMP26 (NM_021801) (SEQ ID NO:63 or 64), ADAM10 (NM_001110) (SEQ ID NO:65 or 66), ADAM8 (NM_001109) (SEQ ID NO:5 or 6), ADAM1 (XM_132370) (SEQ ID NO:67 or 68),
 5 TIM1 (NM_003254) (SEQ ID NO:69 or 70), MUC1 (XM_053256) (SEQ ID NO:71 or 72), CEA (NM_004363) (SEQ ID NO:73 or 74), NCA (NM_002483) (SEQ ID NO:75 or 76), Follistatin (NM_006350) (SEQ ID NO:77 or 78), Claudin 1 (NM_021101) (SEQ ID NO:79 or 80), Claudin 14 (NM_012130) (SEQ ID NO:81 or 82), tenascin-R (NM_003285) (SEQ ID NO:83 or 84), CAD3 (NM_001793) (SEQ ID NO:85 or 86), AXO1 (NM_005076) (SEQ ID
 10 NO:9 or 10), CONT (NM_001843) (SEQ ID NO:87 or 88), Osteopontin (NM_000582) (SEQ ID NO:89 or 90), Galectin 8 (NM_006499) (SEQ ID NO:91 or 92), PGS1 (bilycan, NM_001711) (SEQ ID NO:93 or 94), Frizzled 2 (NM_001466) (SEQ ID NO:95 or 96), ISLR (NM_005545) (SEQ ID NO:97 or 98), FLJ23399 (NM_022763) (SEQ ID NO:99 or 100), TEM1 (NM_020404) (SEQ ID NO:101 or 102), Tie2 ligand2 (NM_001147) (SEQ ID NO:103
 15 or 104), STC-2 (NM_003714) (SEQ ID NO:19 or 20), VEGFC (NM_005429) (SEQ ID NO:105 or 106), tPA (NM_000930) (SEQ ID NO:107 or 108), Endothelin 1 (NM_001955) (SEQ ID NO:1 or 2), Thrombomodulin (NM_000361) (SEQ ID NO:109 or 110), TF (NM_001993) (SEQ ID NO:111 or 112), GPR4 (NM_005282) (SEQ ID NO:113 or 114), GPR66 (NM_006056) (SEQ ID NO:115 or 116), SLC22A2 (NM_003058) ((SEQ ID NO:117
 20 or 118), MLSN1 (NM_002420) (SEQ ID NO:119 or 120), and ATN2 (Na/K transport, NM_000702) (SEQ ID NO:121 or 122), wherein gene expression of at least one, at least two, at least three, at least five, or at least eight genes selected from the group are expressed at a level at least about 1.5-fold above the level in normal control tissue. Where the candidate compound is an antibody, the antibody is alternatively a polyclonal, monoclonal, humanized
 25 antibody, a Fab, a F(ab')₂, or a binding fragment of any one of these compounds.

In an embodiment, the sequences which are used to determine sequence identity or similarity are selected from the sequences described herein. Optionally, sequence variants are naturally occurring allelic variants, sequence variants or splice variants of these sequences.
 30 Sequence identity is typically calculated using the BLAST algorithm, described in Altschul et al Nucleic Acids Res. 25,3389-3402 (1997) with the BLOSUM62 default matrix.

In one embodiment, nucleic acid homology can be determined through hybridisation studies. Nucleic acids which hybridise under stringent conditions to the nucleic acids of the

invention are considered high-grade esophageal dysplasia sequences. Under stringent conditions, hybridisation will most preferably occur at 42°C in 750 mM NaCl, 75 mM trisodium citrate, 2% SDS, 50% formamide, 1X Denhart's, 10% (w/v) dextran sulphate and 100 pg/ml denatured salmon sperm DNA. Useful variations on these conditions will be readily apparent to those skilled in the art. The washing steps which follow hybridization most preferably occur at 65°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

As a result of the degeneracy of the genetic code, a number of polynucleotide sequences encoding polypeptides of the invention, some that may have minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention includes each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring high-grade esophageal dysplasia sequences, and all such variations are to be considered as being specifically disclosed.

The polynucleotides of this invention include RNA, cDNA, genomic DNA, synthetic forms, and mixed polymers, both sense and antisense strands, and may be chemically or biochemically modified, or may contain non-natural or derivatised nucleotide bases as will be appreciated by those skilled in the art. Such modifications include labels, methylation, intercalators, alkylators and modified linkages. In some instances it may be advantageous to produce nucleotide sequences encoding high-grade esophageal dysplasia sequences of the invention, or their derivatives, possessing a substantially different codon usage than that of the naturally occurring gene. For example, codons may be selected to increase the rate of expression of the peptide in a particular prokaryotic or eukaryotic host corresponding with the frequency that particular codons are utilized by the host. Other reasons to alter the nucleotide sequence encoding high-grade esophageal dysplasia sequences of the invention, or their derivatives, without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

In some instances, useful nucleic acid sequences up-regulated in high-grade esophageal dysplasia of the invention are fragments of larger genes and may be used to identify and obtain

corresponding full-length genes. Full-length sequences of the genes selected from the HGD gene marker group or cancer gene marker group of the invention can be obtained using a partial gene sequence using methods known per se to those skilled in the art. For example, "restriction-site PCR" may be used to retrieve unknown sequence adjacent to a portion of DNA whose sequence is known. In this technique universal primers are used to retrieve unknown sequence. Inverse PCR may also be used, in which primers based on the known sequence are designed to amplify adjacent unknown sequences. These upstream sequences may include promoters and regulatory elements. In addition, various other PCR-based techniques may be used, for example a kit available from Clontech (Palo Alto, California) allows for a walking PCR technique, the 5'RACE kit (Gibco-BRL) allows isolation of additional sequence while additional 3' sequence can be obtained using practised techniques.

The present invention allows for the preparation of purified high-grade dysplasia polypeptide (i.e. a polypeptide encoded by a gene disclosed herein as up-regulated in high-grade esophageal dysplasia) or protein, from the polynucleotides of the present invention or variants thereof. In order to do this, host cells may be transfected with a nucleic acid molecule as described above. Typically said host cells are transfected with an expression vector comprising a nucleic acid encoding a high-grade esophageal dysplasia protein according to the invention. Cells are cultured under the appropriate conditions to induce or cause expression of the high-grade esophageal dysplasia protein. The conditions appropriate for high-grade esophageal dysplasia protein expression will vary with the choice of the expression vector and the host cell, and will be easily ascertained by one skilled in the art.

A variety of expression vector/host systems may be utilized to contain and express the high-grade dysplasia sequences of the invention and are well known in the art. These include, but are not limited to, microorganisms such as bacteria transformed with plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e. g., baculovirus); or mouse or other animal or human tissue cell systems. In a preferred embodiment the high-grade esophageal dysplasia proteins of the invention are expressed in mammalian cells using various expression vectors including plasmid, cosmid and viral systems such as adenoviral, retroviral or vaccinia virus expression systems. The invention is not limited by the host cell employed.

The polynucleotide sequences, or variants thereof, of the present invention can be stably expressed in cell lines to allow long term production of recombinant proteins in mammalian systems. These sequences can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. The selectable marker confers resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode a protein of the invention may be designed to contain signal sequences which direct secretion of the protein through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, glycosylation, phosphorylation, and acylation. Post-translational cleavage of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells having specific cellular machinery and characteristic mechanisms for post-translational activities (e. g., CHO or HeLa cells), are available from the American Type Culture Collection (ATCC) and may be chosen to ensure the correct modification and processing of the foreign protein.

When large quantities of protein are needed such as for antibody production, vectors which direct high levels of high-grade esophageal dysplasia gene expression may be used such as those containing the T5 or T7 inducible bacteriophage promoter.

The present invention also includes the use of the expression systems described above in generating and isolating fusion proteins which contain important functional domains of the protein. These fusion proteins are used for binding, structural and functional studies as well as for the generation of appropriate antibodies.

In order to express and purify the protein as a fusion protein, the appropriate cDNA sequence is inserted into a vector which contains a nucleotide sequence encoding another peptide (for example, glutathione succinyl transferase). The fusion protein is expressed and recovered from prokaryotic or eukaryotic cells. The fusion protein can then be purified by affinity chromatography based upon the fusion vector sequence. The relevant protein can subsequently be obtained by enzymatic cleavage of the fusion protein.

In one embodiment, a fusion protein may be generated by the fusion of a high-grade dysplasia polypeptide with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino-or carboxy-terminus of the high-grade esophageal dysplasia polypeptide. The presence of such epitope-tagged forms of a high-grade esophageal dysplasia polypeptide can be detected using an antibody against the tag polypeptide. Also, provision of the epitope tag enables the high-grade dysplasia polypeptide to be readily purified by affinity purification using an anti-tag antibody or another type of affinity matrix that binds to the epitope tag.

Various tag polypeptides and their respective antibodies are well known in the art. Examples include poly-histidine or poly-histidine-glycine tags and the c- myc tag and antibodies thereto. Fragments of high-grade dysplasia polypeptide may also be produced by direct peptide synthesis using solid-phase techniques. Automated synthesis may be achieved by using the ABI 433A Peptide Synthesizer (Applied Biosystems, Foster City, CA). Various fragments of high-grade dysplasia polypeptide may be synthesized separately and then combined to produce the full-length molecule.

In a further aspect of the invention there is provided a method of preparing a polypeptide as described above, comprising the steps of: (1) culturing the host cells under conditions effective for production of the polypeptide; and (2) harvesting the polypeptide.

Substantially purified high-grade dysplasia polypeptide or fragments thereof can then be used in further biochemical analyses to establish secondary and tertiary structure for example by x-ray crystallography of the protein or by nuclear magnetic resonance (NMR). Determination of structure allows for the rational design of pharmaceuticals to interact with the protein, alter protein charge configuration or charge interaction with other proteins, or to alter its function in the cell.

With the identification of the high-grade esophageal dysplasia marker gene nucleotide sequences and the polypeptide sequences encoded by them, probes and antibodies raised to the genes can be used in a variety of hybridisation and immunological assays to screen for and
5 detect the presence of either a normal or mutated gene or gene product.

In addition the nucleotide and protein sequences of the high-grade dysplasia genes provided in this invention enable therapeutic methods for the treatment of cancer, such as adenocarcinoma associated with one or more of these genes, enable screening of compounds
10 for therapeutic intervention, and also enable methods for the diagnosis or prognosis of cancer associated with the these genes. Examples of such cancers include, but are not limited to, esophageal adenocarcinoma.

Transducing retroviral vectors are often used for producing a cell line expressing a
15 gene above the level of expression in a cell lacking the additional copy of the gene. Such a cell is useful according to the invention for the production of a cell line useful for screening candidate compounds capable of reducing expression of a gene associated with high-grade esophageal dysplasia, reducing expression of a polypeptide encoded by the gene, or inhibiting activity of the polypeptide, such that the cell does not progress from dysplasia to cancer. The
20 full-length high-grade dysplasia gene, or portions thereof, can be cloned into a retroviral vector and expression can be driven from its endogenous promoter or from the retroviral long terminal repeat or from a promoter specific for the target cell type of interest. Other viral vectors can be used and include, as is known in the art, adenoviruses, adeno-associated virus, vaccinia virus, papovaviruses, lentiviruses and retroviruses of avian, murine and human origin.

The viral vector described herein above is also useful for gene therapy to reduce the activity of the high-grade dysplasia genes of the invention, such as by antisense expression inhibition or RNA interference (see, for example, Paddison, P.J. et al., *Genes & Development* 16:948-958 (2002) and Brummelkamp, T.R. et al., *Science* 296:550-553 (2002)). Gene
30 therapy would be carried out according to established methods (Friedman, 1991; Culver, 1996). A vector containing a copy of a high-grade esophageal dysplasia gene linked to expression control elements and capable of replicating inside the cells is prepared. Alternatively the vector may be replication deficient and may require helper cells or helper virus for replication and virus production and use in gene therapy.

Gene transfer using non-viral methods of infection can also be used. These methods include direct injection of DNA, uptake of naked DNA in the presence of calcium phosphate, electroporation, protoplast fusion or liposome delivery. Gene transfer can also be achieved by delivery as a part of a human artificial chromosome or receptor-mediated gene transfer. This involves linking the DNA to a targeting molecule that will bind to specific cell-surface receptors to induce endocytosis and transfer of the DNA into mammalian cells. One such technique uses poly-L-lysine to link asialoglycoprotein to DNA. An adenovirus is also added to the complex to disrupt the lysosomes and thus allow the DNA to avoid degradation and move to the nucleus. Infusion of these particles intravenously has resulted in gene transfer into hepatocytes.

Inhibiting high-grade esophageal dysplasia gene or polypeptide function that are up-regulated in cancer can be achieved in a variety of ways as would be appreciated by those skilled in the art. Typically, a vector expressing the complement of a polynucleotide encoding a high-grade dysplasia gene of the invention may be administered to a subject to treat or prevent a disorder associated with increased activity and/or expression of the gene including, but not limited to, those described above.

Antisense strategies may use a variety of approaches including the use of antisense oligonucleotides, ribozymes, DNazymes, injection of antisense RNA and transfection of antisense RNA expression vectors. Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (see, for example, Goldman, CK. et al., Nature Biotechnology 15: 462-466 (1997))

Where purified protein or polypeptide is used to produce antibodies which specifically bind a high-grade dysplasia protein, the antibody(ies) are used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues that express the protein. Such antibodies may include, but are not limited to,

polyclonal, monoclonal, chimeric and single chain antibodies as would be understood by the person skilled in the art.

For the production of antibodies, various hosts including rabbits, rats, goats, mice,
5 humans, and others may be immunized by injection with a protein of the invention or with any fragment or oligopeptide thereof, which has immunogenic properties. Various adjuvants may be used to increase immunological response and include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface-active substances such as lysolecithin. Adjuvants used in humans include BCG (bacilli Calmette-Guerin) and *Corynebacterium*
10 *parvum*.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to the high-grade dysplasia of the invention have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also
15 preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of amino acids from these proteins may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

20

Monoclonal antibodies to high-grade dysplasia polypeptides or proteins of the invention may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma
25 technique. (For example, see Kohler, G. and Milstein, C., *Nature* 256:495-497 (1975); Kozbor, D. et al., *Immunol. Methods* 81:31-42 (1985); and Cole, S.P. et al., *Mol. Cell Biol.* 62:109-120 (1984)).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte
30 population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature.

Antibody fragments which contain specific binding sites for the high-grade esophageal dysplasia proteins may also be generated. For example, such fragments include fragments

produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(AB)₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (For example, see Huse, W. D. et al., Science 246:1275-1281 (1989)).

5 Various immunoassays well known in art may be used for screening to identify antibodies having the desired specificity.

Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art.
10 Such immunoassays typically involve the measurement of complex formation between a protein and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may also be employed.

15 Candidate pharmaceutical agents or compounds encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having molecular weight of more than 100 and less than about 2,500 daltons. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids and steroids and peptides.

20 Agent screening techniques include, but are not limited to, utilising eukaryotic or prokaryotic host cells that are stably transformed with recombinant molecules expressing a particular high-grade dysplasia polypeptide of the invention, or fragment thereof, preferably in competitive binding assays. Binding assays will measure for the formation of complexes
25 between the high-grade esophageal dysplasia polypeptide, or fragments thereof, and the agent being tested, or will measure the degree to which an agent being tested will interfere with the formation of a complex between the high-grade esophageal dysplasia polypeptide, or fragment thereof, and a known ligand.

30 Another technique for drug screening provides high- throughput screening for compounds having suitable binding affinity to a high-grade dysplasia polypeptide. In such a technique, large numbers of small peptide test compounds are synthesised on a solid substrate and can be assayed through high-grade esophageal dysplasia polypeptide binding and washing. Bound high-grade dysplasia polypeptide is then detected by methods well known in

the art. In a variation of this technique, purified polypeptides can be coated directly onto plates to identify interacting test compounds.

5 An additional method for drug screening involves the use of host eukaryotic cell lines which carry mutations in a particular high-grade dysplasia gene. The host cell lines are also defective at the polypeptide level. Other cell lines may be used where the gene expression of the high-grade esophageal dysplasia gene can be switched off or up-regulated. The host cell lines or cells are grown in the presence of various drug compounds and the rate of growth of the host cells is measured to determine if the compound is capable of regulating the growth of
10 defective cells.

A high-grade esophageal dysplasia polypeptide encoded by an HGD marker gene may also be used for screening compounds developed as a result of combinatorial library technology. This provides a way to test a large number of different substances for their ability
15 to modulate activity of a polypeptide. The use of peptide libraries is preferred with such libraries and their use known in the art.

A substance identified as a modulator of polypeptide function may be peptide or non-peptide in nature. Non-peptide "small molecules" are often preferred for many *in vivo*
20 pharmaceutical applications. In addition, a mimic or mimetic of the substance may be designed for pharmaceutical use. The design of mimetics based on a known pharmaceutically active compound (i.e., a "lead compound") is a common approach to the development of novel pharmaceuticals. This is often desirable where the original active compound is difficult or expensive to synthesise or where it provides an unsuitable method of administration. In the
25 design of a mimetic, particular parts of the original active compound that are important in determining the target property are identified. These parts or residues constituting the active region of the compound are known as its pharmacophore. Once found, the pharmacophore structure is modelled according to its physical properties using data from a range of sources including x-ray diffraction data and NMR. A template molecule is then selected onto which
30 chemical groups which mimic the pharmacophore can be added. The selection can be made such that the mimetic is easy to synthesise, is likely to be pharmacologically acceptable, does not degrade *in vivo* and retains the biological activity of the lead compound. Further optimisation or modification can be carried out to select one or more final mimetics useful for *in vivo* or clinical testing.

It is also possible to isolate a target-specific antibody and then solve its crystal structure. In principle, this approach yields a pharmacophore upon which subsequent drug design can be based as described above. It may be possible to avoid protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional, pharmacologically active antibody.

As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analogue of the original binding site. The anti-id could then be used to isolate peptides from chemically or biologically produced peptide banks.

In further embodiments, any of the genes, proteins, antagonists, antibodies, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents.

Selection of the appropriate agents may be made by those skilled in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, therapeutic efficacy with lower dosages of each agent may be possible, thus reducing the potential for adverse side effects.

In a further aspect a pharmaceutical composition and a pharmaceutically acceptable carrier may be administered to a patient diagnosed as experiencing high-grade esophageal dysplasia for the inhibition or prevention of progression of the disease to adenocarcinoma.

The pharmaceutical composition may comprise any one or more of a polypeptide as described above, typically a substantially purified high-grade esophageal dysplasia polypeptide, an antibody to a high-grade esophageal dysplasia polypeptide, a vector capable of expressing a high-grade esophageal dysplasia polypeptide, a compound which increases or decreases expression of a high-grade esophageal dysplasia gene, a candidate drug that restores wild-type activity to a high-grade esophageal dysplasia gene or an antagonist of a high-grade esophageal dysplasia gene.

The pharmaceutical composition may be administered to a subject to treat or prevent a cancer associated with decreased activity and/or expression of a high-grade esophageal dysplasia gene including, but not limited to, those provided above.

- 5 Pharmaceutical compositions in accordance with the present invention are prepared by mixing a polypeptide of the invention, or active fragments or variants thereof, having the desired degree of purity, with acceptable carriers, excipients, or stabilizers which are well known.

10 Acceptable carriers, excipients or stabilizers are nontoxic at the dosages and concentrations employed, and include buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid; low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose,
15 mannose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as Tween, Pluronic or polyethylene glycol (PEG).

20 Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

25 Polynucleotide sequences encoding the high-grade esophageal dysplasia genes of the invention may be used for the diagnosis or prognosis of cancers associated with their dysfunction, or a predisposition to such cancers. Examples of such cancers include, but are not limited to, adenocarcinoma, such as in patients having Barrett's esophagus. Diagnosis or prognosis may be used to determine the severity, type or stage of the disease state in order to initiate an appropriate therapeutic intervention.

30 In another embodiment of the invention, the polynucleotides that may be used for diagnostic or prognostic purposes include oligonucleotide sequences, genomic DNA and complementary RNA and DNA molecules. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which mutations or abnormal expression of the relevant high-grade esophageal dysplasia gene may be correlated with disease. Genomic

DNA used for the diagnosis or prognosis may be obtained from body cells, such as those present in the blood, tissue biopsy, surgical specimen, or autopsy material. The DNA may be isolated and used directly for detection of a specific sequence or may be amplified by the polymerase chain reaction (PCR) prior to analysis. Similarly, RNA or cDNA may also be used, with or without PCR amplification. To detect a specific nucleic acid sequence, direct nucleotide sequencing, reverse transcriptase PCR (RT-PCR), hybridization using specific oligonucleotides, restriction enzyme digest and mapping, PCR mapping, RNase protection, and various other methods may be employed.

Oligonucleotides specific to particular sequences can be chemically synthesized and labelled radioactively or non- radioactively and hybridised to individual samples immobilized on membranes or other solid-supports or in solution. The presence, absence or excess expression of a particular high-grade esophageal dysplasia gene may then be visualized using methods such as autoradiography, fluorometry, or colorimetry.

In a particular aspect, the nucleotide sequences encoding a high-grade esophageal dysplasia gene of the invention may be useful in assays that detect the presence of associated disorders, particularly those mentioned previously. The nucleotide sequences encoding the relevant high-grade esophageal dysplasia gene may be labelled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes.

After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding the high-grade esophageal dysplasia gene in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis or prognosis of a disorder associated with a mutation in a particular high-grade esophageal dysplasia gene of the invention, the nucleotide sequence of the relevant gene can be compared between normal tissue and diseased tissue in order to establish whether the patient expresses a mutant gene.

In order to provide a basis for the diagnosis or prognosis of a disorder associated with abnormal expression of a particular high-grade esophageal dysplasia gene of the invention, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding the relevant high-grade esophageal dysplasia gene, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used.

Another method to identify a normal or standard profile for expression of a particular high-grade esophageal dysplasia gene is through quantitative RT-PCR studies. RNA isolated from body cells of a normal individual, particularly RNA isolated from tumour cells, is reverse transcribed and real-time PCR using oligonucleotides specific for the relevant high-grade esophageal dysplasia gene is conducted to establish a normal level of expression of the gene.

Standard values obtained in both these examples may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays or quantitative RT-PCR studies may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding a particular high-grade esophageal dysplasia gene, or closely related molecules, may be used to identify nucleic acid sequences which encode the gene. The specificity of the probe, whether it is made from a highly specific region, e. g., the 5'regulatory region, or from a less specific region, *e.g.*, a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding the high-grade esophageal dysplasia gene, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the high-grade esophageal dysplasia encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of HGD marker genes disclosed in Table 4 or from genomic sequences including promoters, enhancers, and introns of the genes.

Means for producing specific hybridization probes for DNAs encoding the high-grade esophageal dysplasia genes of the invention include the cloning of polynucleotide sequences encoding these genes or their derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, and are commercially available. Hybridization probes may be labelled by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, or other methods known in the art.

According to a further aspect of the invention there is provided the use of a polypeptide as described above in the diagnosis or prognosis of a cancer associated with a high-grade esophageal dysplasia gene of the invention, or a predisposition to such cancers.

When a diagnostic or prognostic assay is to be based upon a high-grade esophageal dysplasia protein, a variety of approaches are possible. For example, diagnosis or prognosis can be achieved by monitoring differences in the electrophoretic mobility of normal and mutant proteins. Such an approach will be particularly useful in identifying mutants in which charge substitutions are present, or in which insertions, deletions or substitutions have resulted in a significant change in the electrophoretic migration of the resultant protein. Alternatively, diagnosis may be based upon differences in the proteolytic cleavage patterns of normal and mutant proteins, differences in molar ratios of the various amino acid residues, or by functional assays demonstrating altered function of the gene products.

In another aspect, antibodies that specifically bind a high-grade esophageal dysplasia gene of the invention may be used for the diagnosis or prognosis of cancers characterized by abnormal expression of the gene, or in assays to monitor patients being treated with the gene or agonists, antagonists, or inhibitors of the gene. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic or

prognostic assays include methods that utilize the antibody and a label to detect a high-grade esophageal dysplasia gene of the invention in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labelled by covalent or non-covalent attachment of a reporter molecule.

5

A variety of protocols for measuring a high-grade esophageal dysplasia gene of the invention, including ELISA, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of their expression. Normal or standard values for their expression are established by combining body fluids or cell extracts taken from normal
10 mammalian subjects, preferably human, with antibody to the high-grade esophageal dysplasia protein under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of any of the high-grade esophageal dysplasia genes expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation
15 between standard and subject values establishes the parameters for diagnosing disease.

Once an individual has been diagnosed with a cancer, effective treatments can be initiated. These may include administering a selective agonist to the relevant mutant high-grade esophageal dysplasia gene so as to restore its function to a normal level or introduction
20 of the wild-type gene, particularly through gene therapy approaches as described above. Typically, a vector capable of expressing the appropriate full-length high-grade esophageal dysplasia gene or a fragment or derivative thereof may be administered. In an alternative approach to therapy, a substantially purified high-grade esophageal dysplasia polypeptide and a pharmaceutically acceptable carrier may be administered, as described above, or drugs
25 which can replace the function of or mimic the action of the relevant high-grade esophageal dysplasia gene may be administered.

In the treatment of cancers associated with increased high-grade esophageal dysplasia gene expression and/or activity, the affected individual may be treated with a selective
30 antagonist such as an antibody to the relevant protein or an antisense (complement) probe to the corresponding gene as described above, or through the use of drugs which may block the action of the relevant high-grade esophageal dysplasia gene.

In further embodiments, complete cDNAs, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to detect or prognose a disorder, and to develop and monitor the activities of therapeutic agents. Microarrays may be prepared, used, and analyzed using methods known in the art (for example, see Schena, M. et al. PNAS USA 93:10614-10619 (1996); Heller, R.A. et al., PNAS USA 94:2150-2155 (1997); and Heller, M.J., Annual Review of Biomedical Engineering 4:129-53 (2002)).

The present invention also provides for the production of genetically modified (knock-out, knock-down, knock-in and transgenic), non-human animal models transformed with the DNA molecules of the invention. These animals are useful for the study of high-grade esophageal dysplasia gene function, to study the mechanisms of cancer as related to the high-grade esophageal dysplasia genes, for the screening of candidate pharmaceutical compounds, for the creation of explanted mammalian cell cultures which express the protein or mutant protein and for the evaluation of potential therapeutic interventions.

One of the high-grade esophageal dysplasia genes of the invention may have been inactivated by knock-out deletion, and knock-out genetically modified non-human animals are therefore provided.

Animal species which are suitable for use in the animal models of the present invention include, but are not limited to, rats, mice, hamsters, guinea pigs, rabbits, dogs, cats, goats, sheep, pigs, and non-human primates such as monkeys and chimpanzees. For initial studies, genetically modified mice and rats are highly desirable due to their relative ease of maintenance and shorter life spans. For certain studies, transgenic yeast or invertebrates may be suitable and preferred because they allow for rapid screening and provide for much easier handling. For longer term studies, non-human primates may be desired due to their similarity with humans.

To create an animal model for a mutated high-grade esophageal dysplasia gene of the invention several methods can be employed. These include generation of a specific mutation

in a homologous animal gene, insertion of a wild type human gene and/or a humanized animal gene by homologous recombination, insertion of a mutant (single or multiple) human gene as genomic or minigene cDNA constructs using wild type or mutant or artificial promoter elements or insertion of artificially modified fragments of the endogenous gene by homologous recombination. The modifications include insertion of mutant stop codons, the deletion of DNA sequences, or the inclusion of recombination elements (lox p sites) recognized by enzymes such as Cre recombinase.

To create a transgenic mouse, which is preferred, a mutant version of a particular high-grade esophageal dysplasia gene of the invention can be inserted into a mouse germ line using standard techniques of oocyte microinjection or transfection or microinjection into embryonic stem cells. Alternatively, if it is desired to inactivate or replace the endogenous high-grade esophageal dysplasia gene, homologous recombination using embryonic stem cells may be applied. For oocyte injection, one or more copies of the mutant or wild type high-grade esophageal dysplasia gene can be inserted into the pronucleus of a just-fertilized mouse oocyte. This oocyte is then reimplanted into a pseudo-pregnant foster mother. The liveborn mice can then be screened for integrants using analysis of tail DNA for the presence of human high-grade esophageal dysplasia gene sequences. The transgene can be either a complete genomic sequence injected as a YAC, BAC, PAC or other chromosome DNA fragment, a cDNA with either the natural promoter or a heterologous promoter, or a minigene containing all of the coding region and other elements found to be necessary for optimum expression. The genetically modified non-human animals as described above are useful for the screening of candidate pharmaceutical compounds.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B are graphs showing a distribution of expression of IL-1H1 (Fig. 1A) and CYP2J2 (Fig. 1B) in the dysplasia-carcinoma sequence in BE. Expression in normal epithelium and in esophageal epithelia from samples of Barrett's esophagus (BE), dysplasia (D), BE adjacent to adenocarcinoma (BE-CA); and adenocarcinoma (CA) are plotted. The vertical line denotes the average Z score in each disease group. Normal refers to the normal esophagus group. Dysplasia includes low- and high-grade dysplasia samples.

Figures 2A and 2B are graphs showing a distribution of expression of AGR2 (Fig. 2A) and NROB2 (Fig. 2B) in the dysplasia-carcinoma sequence in BE. Expression in esophageal epithelia from samples of Barrett's esophagus (BE), dysplasia (D), BE adjacent to adenocarcinoma (BE-CA); and adenocarcinoma (CA) are plotted. The vertical line denotes the average Z score in each disease group. Normal refers to pooled epithelia samples. Dysplasia includes low- and high-grade dysplasia samples.

Figures 3A and 3B are graphs showing a distribution of expression of TCF4 (Fig. 3A) and FLJ23399 (Fig. 3B) in the dysplasia-carcinoma sequence in BE. Expression in esophageal epithelia from samples of Barrett's esophagus (BE), dysplasia (D), BE adjacent to adenocarcinoma (BE-CA); and adenocarcinoma (CA) are plotted. The vertical line denotes the average Z score in each disease group. Normal refers to pooled epithelia samples. Dysplasia includes low- and high-grade dysplasia samples.

Figures 4A and 4B show the nucleic acid sequence (SEQ ID NO:1) and the amino acid sequence (SEQ ID NO:2) of ET-1 (endothelin-1, NM_001955).

Figures 5A and 5B show the nucleic acid sequence (SEQ ID NO:3) and the amino acid sequence (SEQ ID NO:4) of AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408).

Figures 6A and 6B show the nucleic acid sequence (SEQ ID NO:5) and the amino acid sequence (SEQ ID NO:6) of ADAM8 (NM_001109).

Figures 7A and 7B show the nucleic acid sequence (SEQ ID NO:7) and the amino acid sequence (SEQ ID NO:8) of PSS8 (Prostasin precursor, serine protease, NM_002773).

Figures 8A-8C show the nucleic acid sequence (SEQ ID NO:9) and Figure 8D shows the amino acid sequence (SEQ ID NO:10) of AXO1 (Axonin-1 precursor, NM_005076).

Figures 9A and 9B show the nucleic acid sequence (SEQ ID NO:11) and the amino acid sequence (SEQ ID NO:12) of NROB2 (Nuclear hormone receptor, NM_021969).

Figures 10A and 10B show the nucleic acid sequence (SEQ ID NO:13) and the amino acid sequence (SEQ ID NO:14) of TM7SF1 (NM_003272).

Figures 11A and 11B show the nucleic acid sequence (SEQ ID NO:15) and the amino acid sequence (SEQ ID NO:16) of DLDH (dihydrolipamide dehydrogenase, NM_000108).

Figures 12A and 12B show the nucleic acid sequence (SEQ ID NO:17) and the amino acid sequence (SEQ ID NO:18) of MAT2B (methionine adenosyltransferase II, beta, NM_013283).

Figures 13A and 13B show the nucleic acid sequence (SEQ ID NO:19) and the amino acid sequence (SEQ ID NO:20) of STC-2 (stanniocalcin-2, NM_003714).

Figures 14A and 14B show the nucleic acid sequence (SEQ ID NO:21) and the amino acid sequence (SEQ ID NO:22) of PPBI (alkaline phosphatase, intestinal precursor, NM_001631).

Figures 15A and 15B show the nucleic acid sequence (SEQ ID NO:23) and the amino acid sequence (SEQ ID NO:24) of SLNAC1 (sodium channel receptor SLNAC1, NM_004769).

Figures 16A and 16B show the nucleic acid sequence (SEQ ID NO:25) and the amino acid sequence (SEQ ID NO:26) of CAH4 (carbonic anhydrase iv precursor, NM_000717).

Figures 17A and 17B show shows the nucleic acid sequence (SEQ ID NO:27) and the amino acid sequence (SEQ ID NO:28) of PA21 (phopholipase a2 precursor, NM_000928).

Figures 18A and 18B show the nucleic acid sequence (SEQ ID NO: 29) and the amino acid sequence (SEQ ID NO:30) of PAR2 (proteinase activated receptor 2 precursor, NM_005242).

Figures 19A and 19B show the nucleic acid sequence (SEQ ID NO:31) and the amino acid sequence (SEQ ID NO:32) of IDE (insulin-degrading enzyme, NM_004969).

Figures 20A-20B show the nucleic acid sequence (SEQ ID NO:33) and Figure 20C shows the amino acid sequence (SEQ ID NO:34) of MYO1A (myosin-1A, NM_005379).

Figures 21A and 21B the nucleic acid sequence (SEQ ID NO:35) and the amino acid sequence (SEQ ID NO:36) of CYP2J2 (cytochrome P450 monooxygenase, NM_000775).

Figures 22A and 22B show the nucleic acid sequence (SEQ ID NO:37) and the amino acid sequence (SEQ ID NO:38) of PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214).

Figures 23A and 23B show the nucleic acid sequence (SEQ ID NO:39) and the amino acid sequence (SEQ ID NO:40) of CYB5 (cytochrome b5, 3' end, NM_001914).

Figures 24A and 24B show the nucleic acid sequence (SEQ ID NO:41) and the amino acid sequence (SEQ ID NO:42) of COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863).

Figures 25A and 25B show the nucleic acid sequence (SEQ ID NO:43) and the amino acid sequence (SEQ ID NO:44) of TCF4 (NM_030756).

Figures 26A-26B show the nucleic acid sequence (SEQ ID NO:45) and Figure 26C shows the amino acid sequence (SEQ ID NO:46) of CAD17 (liver-intestine cadherin, NM_004063).

Figures 27A and 27B show the nucleic acid sequence (SEQ ID NO:47) and the amino acid sequence (SEQ ID NO:48) of CLDN15 (claudin 15, NM_014343).

Figures 28A-28B show the nucleic acid sequence (SEQ ID NO:49) and Figure 28C shows the amino acid sequence (SEQ ID NO:50) of CFTR (chloride channel, NM_000492).

Figures 29A and 29B show the nucleic acid sequence (SEQ ID NO:51) and the amino acid sequence (SEQ ID NO:52) of H2R (histamine H2 receptor, NM_022304).

Figures 30A-30B show the nucleic acid sequence (SEQ ID NO:53) and Figure 30C shows the amino acid sequence (SEQ ID NO:54) of EGFR (epidermal growth factor receptor, NM_005228).

Figures 31A-31B show the nucleic acid sequence (SEQ ID NO:55) and Figure 31C shows the amino acid sequence (SEQ ID NO:56) of EPHB2, NM_004442).

5 Figures 32A and 32B show the nucleic acid sequence (SEQ ID NO:57) and the amino acid sequence (SEQ ID NO:58) of CRIPTO CR-1 (NM_003212).

Figures 33A and 33B show the nucleic acid sequence (SEQ ID NO:59) and the amino acid sequence (SEQ ID NO:60) of Eprin B1 (NM_004429).

10 Figures 34A and 34B show the nucleic acid sequence (SEQ ID NO:61) and the amino acid sequence (SEQ ID NO:62) of MMP-17/MT4-MMP (matrix metalloproteinase 17, NM_016155).

15 Figures 35A and 35B show the the nucleic acid sequence (SEQ ID NO:63) and the amino acid sequence (SEQ ID NO:64) of MMP26 (matrix metalloproteinase 26, NM_021801).

Figures 36A and 36B show the nucleic acid sequence (SEQ ID NO:65) and the amino acid sequence (SEQ ID NO:66) of ADAM10 (NM_001110).

20 Figures 37A and 37B show the nucleic acid sequence (SEQ ID NO:67) and the amino acid sequence (SEQ ID NO:68) of ADAM1 (XM_132370).

Figures 38A and 38B show the nucleic acid sequence (SEQ ID NO:69) and the amino acid sequence (SEQ ID NO:70) of TIM1(NM_003254).

25 Figures 39A and 39B show the nucleic acid sequence (SEQ ID NO:71) and the amino acid sequence (SEQ ID NO:72) of MUC1 (XM_053256).

30 Figures 40A and 40B show the nucleic acid sequence (SEQ ID NO:73) and the amino acid sequence (SEQ ID NO:74) of CEA (NM_004363).

Figures 41A and 41B show the nucleic acid sequence (SEQ ID NO:75) and the amino acid sequence (SEQ ID NO:76) of NCA (NM_002483).

Figures 42A and 42B show the nucleic acid sequence (SEQ ID NO:77) and the amino acid sequence (SEQ ID NO:78) of Follistatin (NM_006350).

5 Figures 43A and 43B show the nucleic acid sequence (SEQ ID NO:79) and the amino acid sequence (SEQ ID NO:80) of Claudin 1 (NM_021101).

Figures 44A and 44B show the nucleic acid sequence (SEQ ID NO:81) and the amino acid sequence (SEQ ID NO:82) of Claudin 14 (NM_012130).

10

Figures 45A-45B show the nucleic acid sequence (SEQ ID NO:83) and Figure 45C show the amino acid sequence (SEQ ID NO:84) of Tenascin-R (NM-003285).

15 Figures 46A and 46B show the nucleic acid sequence (SEQ ID NO:85) and the amino acid sequence (SEQ ID NO:86) of CAD3 (NM_001793).

Figures 47A and 47B show the nucleic acid sequence (SEQ ID NO:87) and the amino acid sequence (SEQ ID NO:88) of CONT (NM_001843).

20 Figures 48A and 48B show the nucleic acid sequence (SEQ ID NO:89) and the amino acid sequence (SEQ ID NO:90) of Osteopontin (NM_000582).

Figures 49A and 49B show the nucleic acid sequence (SEQ ID NO:91) and the amino acid sequence (SEQ ID NO:92) of Galectin 8 (NM_006499).

25

Figures 50A and 50B show the nucleic acid sequence (SEQ ID NO:93) and the amino acid sequence (SEQ ID NO:94) of GS1 (bihlycan, NM_001711).

30 Figures 51A and 51B show the nucleic acid sequence (SEQ ID NO:95) and the amino acid sequence (SEQ ID NO:96) of Fizzled 2 (NM001466).

Figures 52A and 52B show the nucleic acid sequence (SEQ ID NO:97) and the amino acid sequence (SEQ ID NO:98) of ISLR (NM_005545).

Figures 53A-53B show the nucleic acid sequence (SEQ ID NO:) and Figure 53C shows the amino acid sequence (SEQ ID NO:2) of

Figures 54A and 54B show the nucleic acid sequence (SEQ ID NO:1) and the amino acid sequence (SEQ ID NO:2) of

Figures 55A and 55B show the nucleic acid sequence (SEQ ID NO:103) and the amino acid sequence (SEQ ID NO:104) of Tie2 ligand2 (NM_001147).

Figures 56A and 56B show the nucleic acid sequence (SEQ ID NO:105) and the amino acid sequence (SEQ ID NO:106) of VEGFC (NM_005429).

Figures 57A and 57B show the nucleic acid sequence (SEQ ID NO:107) and the amino acid sequence (SEQ ID NO:108) of tPA (NM_000930).

Figures 58A-58B show the nucleic acid sequence (SEQ ID NO:109) and Figure 58C shows the amino acid sequence (SEQ ID NO:110) of thrombomodulin (NM_000361).

Figures 59A and 59B show the nucleic acid sequence (SEQ ID NO:111) and the amino acid sequence (SEQ ID NO:112) of TF (coagulation factor III, thromboplastin, tissue factor, NM_0001993).

Figures 60A and 60B show the nucleic acid sequence (SEQ ID NO:113) and the amino acid sequence (SEQ ID NO:114) of GPR4 (G-coupled protein receptor-4, NM_005282).

Figures 61A and 61B show the nucleic acid sequence (SEQ ID NO:115) and the amino acid sequence (SEQ ID NO:116) of GPR66 (G-coupled protein receptor 66).

Figures 62A and 62B show the nucleic acid sequence (SEQ ID NO:117) and the amino acid sequence (SEQ ID NO:118) of SLC22A2 (NM_003058).

Figures 63A-63B show the nucleic acid sequence (SEQ ID NO:119) and Figure 63C shows the amino acid sequence (SEQ ID NO:120) of MLSN1 (NM_002420).

Figures 64A-64B show the nucleic acid sequence (SEQ ID NO:121) and Figure 64C shows the amino acid sequence (SEQ ID NO:122) of ATN2 (Na/K transport, NM_000702).

DESCRIPTION OF THE INVENTION

5

Barrett's esophagus, a complication of gastrointestinal reflux disease, is the primary risk factor for esophageal adenocarcinoma. Biopsy specimens representing disease progression through Barrett's esophagus, dysplasia and adenocarcinoma, were collected and analyzed using cDNA microarrays to identify genes expressed in the different disease stages. It was discovered that the expression of particular genes increased with the progression of the disease through dysplasia, especially high grade dysplasia, suggestive of a differentiated small intestinal enterocyte lineage. The present invention defines a collection of markers that assist in identifying patients with highest risk of developing cancer, especially the development of esophageal adenocarcinoma.

15

The progression of Barrett's esophagus through dysplasia to adenocarcinoma was examined, identifying specific genes associated with increasing risk of carcinogenesis. These data provide insight into the potential role of progressive intestinal metaplasia in generating the colon tumor-like expression profiles disclosed herein for esophageal adenocarcinoma. Genes that define early stages of this process, progression of BE to dysplasia, serve as markers to permit targeting of surveillance to those patients at most risk of developing esophageal carcinoma.

20

DNA microarray technology has been used to characterize and cluster Barrett's metaplasia from normal mucosa, and esophageal adenocarcinoma and squamous cell carcinoma (Barrett et al., Neoplasia 4:121-128 (2002); and Selaru et al., Oncogene 21:475-478 (2002)). The authors do not, however, describe HGD markers or dysplasia markers of any kind useful for predicting patients likely to develop adenocarcinoma.

25

The present invention provides nucleic acid and protein sequences that are differentially expressed in high-grade esophageal dysplasia when compared to normal tissue controls, here-in termed "high-grade dysplasia genes," "high-grade dysplasia nucleic acid sequences," "HGD marker genes" and the like. As outlined below, high-grade esophageal dysplasia sequences that are differentially expressed include those that are up-regulated in

30

high-grade esophageal dysplasia). The differential expression of these sequences in high-grade esophageal dysplasia combined with the fact they have been identified in patients likely to develop cancer, such as adenocarcinoma, they are contributory factors in cancer. The high-grade esophageal dysplasia nucleic acid sequences, or the polypeptides encoded by the nucleic acids, of the invention are disclosed in Table 4 as HGD marker genes, or polypeptides, as follows: ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1 or 2); AGR2 (anterior gradient 2 (Xenopus laevis) homolog, NM_006408) (SEQ ID NO:3 or 4); ADAM8 (NM_001109) (SEQ ID NO:5 or 6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7 or 8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9 or 10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11 or 12); TM7SF1 (NM_003272) (SEQ ID NO:13 or 14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15 or 16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17 or 18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19 or 20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21 or 22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23 or 24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25 or 26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27 or 28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29 or 30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31 or 32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33 or 34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35 or 36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37 or 38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39 or 40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41 or 42); and TCF4 (NM_030756) (SEQ ID NO:43 or 44).

Definitions

The phrases "gene amplification" and "gene duplication" are used interchangeably and refer to a process by which multiple copies of a gene or gene fragment are formed in a particular cell or cell line. The duplicated region (a stretch of amplified DNA) is often referred to as "amplicon." Usually, the amount of the messenger RNA (mRNA) produced, *i.e.*, the level of gene expression, also increases in the proportion of the number of copies made of the particular gene expressed.

"Tumor", as used herein, refers to all neoplastic cell growth and proliferation, whether malignant or benign, and all pre-cancerous and cancerous cells and tissues.

The terms "cancer" and "cancerous" refer to or describe the physiological condition in mammals that is typically characterized by unregulated cell growth. Examples of cancer include but are not limited to, carcinoma, adenocarcinoma; lymphoma, blastoma, sarcoma, and leukemia. More particular examples of such cancers include esophageal cancer, breast cancer, prostate cancer, colon cancer, squamous cell cancer, small-cell lung cancer, non-small cell lung cancer, gastrointestinal cancer, pancreatic cancer, glioblastoma, cervical cancer, ovarian cancer, liver cancer, bladder cancer, hepatoma, colorectal cancer, endometrial carcinoma, salivary gland carcinoma, kidney cancer, liver cancer, vulval cancer, thyroid cancer, hepatic carcinoma and various types of head and neck cancer.

The term "diagnosis" or "diagnosing" as used herein shall refer to the determination of the nature of a case of a disease, such as by determining a gene expression profile or polypeptide expression profile unique to the disease or a stage of the disease.

A "normal" tissue sample refers to tissue or cells that are not diseased as defined herein, such as tissue from a mammal that is not experiencing a particular disease of interest. The term "normal cell" or "normal tissue" as used herein refers to a state of a cell or tissue in which the cell or tissue is apparently free of an adverse biological condition when compared to a diseased cell or tissue having that adverse biological condition. The normal cell or normal tissue may be from any prokaryotic or eukaryotic organism including, but not limited to, bacteria, yeast, insect, bird, reptile, and any mammal including human. Where the normal tissue or cell is used as a normal control sample, it is generally from the same species as the test sample. Where the cell or tissue is mammalian, the cell or tissue is any cell or tissue including, but not limited to blood, muscle, nerve, brain, breast, heart, lung, liver, pancreas, spleen, thymus, esophagus, stomach, intestine, kidney, testis, ovary, uterus, hair follicle, skin, bone, bladder, and spinal cord.

"Treatment" is an intervention performed with the intention of preventing the development or altering the pathology of a disorder. Accordingly, "treatment" refers to both therapeutic treatment and prophylactic or preventative measures. Those in need of treatment include those already with the disorder as well as those in which the disorder is to be

prevented. In tumor (*e.g.*, cancer) treatment, a therapeutic agent may directly decrease the pathology of tumor cells, or render the tumor cells more susceptible to treatment by other therapeutic agents, *e.g.*, radiation and/or chemotherapy.

5 A "pharmaceutical composition" as used herein refers to a composition comprising a chemotherapeutic agent for treatment of a disease combined with physiologically acceptable materials such as carriers, excipients, stabilizers, buffers, salts, antioxidants, hydrophilic polymers, amino acids, carbohydrates, ionic or nonionic surfactants, and/or polyethylene or propylene glycol. The pharmaceutical composition may be in aqueous form, tablet, capsule,
10 microcapsules, liposomes, transdermal patches, and the like.

The "pathology" of cancer includes all phenomena that compromise the well-being of the patient. This includes, without limitation, abnormal or uncontrollable cell growth, metastasis, interference with the normal functioning of neighboring cells, release of cytokines
15 or other secretory products at abnormal levels, suppression or aggravation of inflammatory or immunological response, etc.

"Mammal" for purposes of treatment refers to any animal classified as a mammal, including humans, domestic and farm animals, and zoo, sports, or pet animals, such as dogs,
20 horses, cats, cattle, pigs, sheep, etc. Preferably, the mammal is human.

"Carriers" as used herein include pharmaceutically acceptable carriers, excipients, or stabilizers which are nontoxic to the cell or mammal being exposed thereto at the dosages and concentrations employed. Often the physiologically acceptable carrier is an aqueous pH
25 buffered solution. Examples of physiologically acceptable carriers include buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid; low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides,
30 and other carbohydrates including glucose, mannose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as TWEENTM, polyethylene glycol (PEG), and PLURONICSTM.

Administration "in combination with" one or more further therapeutic agents includes simultaneous (concurrent) and consecutive administration in any order.

The term "cytotoxic agent" as used herein refers to a substance that inhibits or prevents the function of cells and/or causes destruction of cells. The term is intended to include radioactive isotopes (*e.g.*, I^{131} , I^{125} , Y^{90} and Re^{186}), chemotherapeutic agents, and toxins such as enzymatically active toxins of bacterial, fungal, plant or animal origin, or fragments thereof.

A "chemotherapeutic agent" is a chemical compound useful in the treatment of cancer. Examples of chemotherapeutic agents include adriamycin, doxorubicin, epirubicin, 5-fluorouracil, cytosine arabinoside ("Ara-C"), cyclophosphamide, thiotepe, busulfan, cytoxin, taxoids, *e.g.*, paclitaxel (Taxol, Bristol-Myers Squibb Oncology, Princeton, NJ), and doxetaxel (Taxotere, Rhône-Poulenc Rorer, Antony, France), taxotere, methotrexate, cisplatin, melphalan, vinblastine, bleomycin, etoposide, ifosfamide, mitomycin C, mitoxantrone, vincristine, vinorelbine, carboplatin, teniposide, daunomycin, carminomycin, aminopterin, dactinomycin, mitomycins, esperamicins (see U.S. Pat. No. 4,675,187), 5-FU, 6-thioguanine, 6-mercaptopurine, actinomycin D, VP-16, chlorambucil, melphalan, and other related nitrogen mustards. Also included in this definition are hormonal agents that act to regulate or inhibit hormone action on tumors such as tamoxifen and onapristone. In an embodiment, the chemotherapeutic agent of the invention is a chemical compound useful in the treatment of HGD, adenocarcinoma, or for inhibiting or preventing progression from the HGD to adenocarcinoma in a patient.

A "growth inhibitory agent" when used herein refers to a compound or composition which inhibits growth of a cell, especially cancer cell overexpressing any of the genes identified herein, either *in vitro* or *in vivo*. Thus, the growth inhibitory agent is one which significantly reduces the percentage of cells overexpressing such genes in S phase. Examples of growth inhibitory agents include agents that block cell cycle progression (at a place other than S phase), such as agents that induce G1 arrest and M-phase arrest. Classical M-phase blockers include the vincas (vincristine and vinblastine), taxol, and topo II inhibitors such as doxorubicin, epirubicin, daunorubicin, etoposide, and bleomycin. Those agents that arrest G1 also spill over into S-phase arrest, for example, DNA alkylating agents such as tamoxifen, prednisone, dacarbazine, mechlorethamine, cisplatin, methotrexate, 5-fluorouracil, and ara-C. Further information can be found in The Molecular Basis of Cancer, Mendelsohn and Israel,

eds., Chapter 1, entitled "Cell cycle regulation, oncogens, and antineoplastic drugs" by Murakami *et al.*, (WB Saunders: Philadelphia, 1995), especially p. 13.

"Doxorubicin" is an anthracycline antibiotic. The full chemical name of doxorubicin is
5 (8S-cis)-10-[(3-amino-2,3,6-trideoxy- α -L-lyxo-hexapyranosyl)oxy]-7,8,9,10-tetrahydro-
6,8,11-trihydroxy-8-(hydroxyacetyl)-1-methoxy-5,12-naphthacenedione.

The term "cytokine" is a generic term for proteins released by one cell population which act on another cell as intercellular mediators. Examples of such cytokines are
10 lymphokines, monokines, and traditional polypeptide hormones. Included among the cytokines are growth hormone such as human growth hormone, N-methionyl human growth hormone, and bovine growth hormone; parathyroid hormone; thyroxine; insulin; proinsulin; relaxin; prorelaxin; glycoprotein hormones such as follicle stimulating hormone (FSH), thyroid stimulating hormone (TSH), and luteinizing hormone (LH); hepatic growth factor;
15 fibroblast growth factor; prolactin; placental lactogen; tumor necrosis factor- α and - β ; mullerian-inhibiting substance; mouse gonadotropin-associated peptide; inhibin; activin; vascular endothelial growth factor; integrin; thrombopoietin (TPO); nerve growth factors such as NGF- β ; platelet-growth factor; transforming growth factors (TGFs) such as TGF- α and TGF- β ; insulin-like growth factor-I and -II; erythropoietin (EPO); osteoinductive factors;
20 interferons such as interferon - α , - β , and - γ ; colony stimulating factors (CSFs) such as macrophage-CSF (M-CSF); granulocyte-macrophage-CSF (GM-CSF); and granulocyte-CSF (G-CSF); interleukins (ILs) such as IL-1, IL-1a, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-11, IL-12; a tumor necrosis factor such as TNF- α or TNF- β ; and other polypeptide factors including LIF and kit ligand (KL). As used herein, the term cytokine includes proteins from
25 natural sources or from recombinant cell culture and biologically active equivalents of the native sequence cytokines.

The term "prodrug" as used in this application refers to a precursor or derivative form of a pharmaceutically active substance that is less cytotoxic to tumor cells compared to the
30 parent drug and is capable of being enzymatically activated or converted into the more active parent form. *See, e.g.*, Wilman, "Prodrugs in Cancer Chemotherapy", Biochemical Society Transactions, 14:375-382, 615th Meeting, Belfast (1986), and Stella *et al.*, "Prodrugs: A Chemical Approach to Targeted Drug Delivery", Directed Drug Delivery, Borchardt *et al.*, (ed.), pp. 147-267, Humana Press (1985). The prodrugs of this invention include, but are not

limited to, phosphate-containing prodrugs, thiophosphate-containing prodrugs, sulfate-containing prodrugs, peptide-containing prodrugs, D-amino acid-modified prodrugs, glycosylated prodrugs, β -lactam-containing prodrugs, optionally substituted phenoxyacetamide-containing prodrugs or optionally substituted phenylacetamide-containing prodrugs, 5-fluorocytosine and other 5-fluorouridine prodrugs which can be converted into the more active cytotoxic free drug. Examples of cytotoxic drugs that can be derivatized into a prodrugs form for use in this invention include, but are not limited to, those chemotherapeutic agents described above.

An “effective amount” or therapeutically effective amount” of a polypeptide disclosed herein or an antagonist thereof, in reference to inhibition of neoplastic cell growth, tumor growth or cancer cell growth, is an amount capable of inhibiting, to some extent, the growth of target cells. The term includes an amount capable of invoking a growth inhibitory, cytostatic and/or cytotoxic effect and/or apoptosis of the target cells. An “effective amount” is an amount of an antagonist of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1 or 2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3 or 4); ADAM8 (NM_001109) (SEQ ID NO:5 or 6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7 or 8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9 or 10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11 or 12); TM7SF1 (NM_003272) (SEQ ID NO:13 or 14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15 or 16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17 or 18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19 or 20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21 or 22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23 or 24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25 or 26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27 or 28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29 or 30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31 or 32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33 or 34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35 or 36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37 or 38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39 or 40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41 or 42); and TCF4 (NM_030756) (SEQ ID NO:43 or 44) gene or polypeptide for purposes of inhibiting neoplastic cell growth, tumor growth or cancer cell growth, may be determined empirically

and in a routine manner. The terms further refer to an amount capable of invoking one or more of the following effects: (1) inhibition, to some extent, of tumor growth, including, slowing down and complete growth arrest; (2) reduction in the number of tumor cells; (3) reduction in tumor size; (4) inhibition (*i.e.*, reduction, slowing down or complete stopping) of tumor cell infiltration into peripheral organs; (5) inhibition (*i.e.*, reduction, slowing down or complete stopping) of metastasis; (6) enhancement of anti-tumor immune response, which may, but does not have to, result in the regression or rejection of the tumor; and/or (7) relief, to some extent, of one or more symptoms associated with the disorder. A “therapeutically effective amount” of an antagonist of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1 or 2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3 or 4); ADAM8 (NM_001109) (SEQ ID NO:5 or 6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7 or 8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9 or 10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11 or 12); TM7SF1 (NM_003272) (SEQ ID NO:13 or 14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15 or 16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17 or 18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19 or 20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21 or 22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23 or 24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25 or 26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27 or 28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29 or 30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31 or 32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33 or 34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35 or 36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37 or 38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39 or 40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41 or 42); or TCF4 (NM_030756) (SEQ ID NO:43 or 44) gene or polypeptide for purposes of treatment of tumor may be determined empirically and in a routine manner.

A “growth inhibitory amount” of a compound that inhibits growth of a cell expressing genes, or polypeptides, from the following group: ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1 or 2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3 or 4); ADAM8 (NM_001109) (SEQ ID NO:5 or 6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7 or 8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ

ID NO:9 or 10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11 or 12); TM7SF1 (NM_003272) (SEQ ID NO:13 or 14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15 or 16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17 or 18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19 or 20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21 or 22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23 or 24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25 or 26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27 or 28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29 or 30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31 or 32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33 or 34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35 or 36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37 or 38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39 or 40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41 or 42); and TCF4 (NM_030756) (SEQ ID NO:43 or 44) is an amount of the compound capable of inhibiting the growth of a cell, especially tumor, *e.g.*, cancer cell, either *in vitro* or *in vivo*. Optionally, the compound is an antagonist of the gene or polypeptide, such as an antagonist antibody or antagonist small organic molecule. A "growth inhibitory amount" of such a compound, for purposes of inhibiting neoplastic cell growth, may be determined empirically and in a routine manner.

A "cytotoxic amount" of an ET-1 (endothelin-1, NM_001955) (SEQ ID NO:2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:4); ADAM8 (NM_001109) (SEQ ID NO:6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:12); TM7SF1 (NM_003272) (SEQ ID NO:14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NO:16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:32); MYO1A (myosin-1A, NM_005379) (SEQ

ID NO:34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:42); or TCF4 (NM_030756) (SEQ ID NO:44) polypeptide antagonist is an amount capable of causing the destruction of a cell, especially tumor, *e.g.*, cancer cell, either *in vitro* or *in vivo*. A "cytotoxic amount" of a such a polypeptide antagonist for purposes of inhibiting neoplastic cell growth may be determined empirically and in a routine manner.

The terms ET-1 (endothelin-1, NM_001955) (SEQ ID NO:2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:4); ADAM8 (NM_001109) (SEQ ID NO:6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:12); TM7SF1 (NM_003272) (SEQ ID NO:14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NO:16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:42); and TCF4 (NM_030756) (SEQ ID NO:44) polypeptide or protein when used herein encompass native sequence ET-1 (endothelin-1, NM_001955) (SEQ ID NO:2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:4); ADAM8 (NM_001109) (SEQ ID NO:6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:12); TM7SF1 (NM_003272) (SEQ ID NO:14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NO:16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:20);

PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:42); and TCF4 (NM_030756) (SEQ ID NO:44) polypeptide variants (which are further defined herein). The ET-1 (endothelin-1, NM_001955) (SEQ ID NO:2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:4); ADAM8 (NM_001109) (SEQ ID NO:6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:12); TM7SF1 (NM_003272) (SEQ ID NO:14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NO:16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:42); or TCF4 (NM_030756) (SEQ ID NO:44) polypeptide may be isolated from a variety of sources, such as from human tissue types or from another source, or prepared by recombinant and/or synthetic methods.

30

A "native sequence polypeptide" of each HGD marker polypeptide has the same amino acid sequence or is a polypeptide variant having at least about 80% amino acid sequence identity, preferably at least about 81% amino acid sequence identity, more preferably at least about 82% amino acid sequence identity, more preferably at least about 83% amino acid

sequence identity, more preferably at least about 84% amino acid sequence identity, more preferably at least about 85% amino acid sequence identity, more preferably at least about 86% amino acid sequence identity, more preferably at least about 87% amino acid sequence identity, more preferably at least about 88% amino acid sequence identity, more preferably at least about 89% amino acid sequence identity, more preferably at least about 90% amino acid sequence identity, more preferably at least about 91% amino acid sequence identity, more preferably at least about 92% amino acid sequence identity, more preferably at least about 93% amino acid sequence identity, more preferably at least about 94% amino acid sequence identity, more preferably at least about 95% amino acid sequence identity, more preferably at least about 96% amino acid sequence identity, more preferably at least about 97% amino acid sequence identity, more preferably at least about 98% amino acid sequence identity and most preferably at least about 99% amino acid sequence identity with a full-length native sequence polypeptide sequence, lacking the signal peptide as disclosed herein, as the ET-1 (endothelin-1, NM_001955) (SEQ ID NO:2); AGR2 (anterior gradient 2 (Xenopus laevis) homolog, NM_006408) (SEQ ID NO:4); ADAM8 (NM_001109) (SEQ ID NO:6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:12); TM7SF1 (NM_003272) (SEQ ID NO:14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NO:16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:42); or TCF4 (NM_030756) (SEQ ID NO:44) polypeptide as derived from nature. Such native sequence polypeptide can be isolated from nature or can be produced by recombinant and/or synthetic means. The term "native sequence polypeptide" specifically encompasses naturally-occurring truncated or secreted forms (*e.g.*, an extracellular domain sequence), naturally-occurring variant forms (*e.g.*, alternatively spliced forms) and

naturally-occurring allelic variants of the polypeptides encoded by a HGD marker gene as disclosed herein. In one embodiment of the invention, the native sequence HGD marker polypeptide is a mature or full-length native sequence HGD marker polypeptide as encoded by the nucleic acid sequences of the GenBank accession numbers listed in Table 4A for the
5 respective polypeptide. Also, the HGD marker polypeptides encoded by the nucleic acid sequences disclosed in the respective GenBank accession numbers listed in Table 4A, are shown to begin with the methionine residue designated therein as amino acid position 1, it is conceivable and possible that another methionine residue located either upstream or downstream from amino acid position 1 may be employed as the starting amino acid residue
10 for HGD marker polypeptide.

The “extracellular domain” or “ECD” of a polypeptide disclosed herein refers to a form of the polypeptide which is essentially free of the transmembrane and cytoplasmic domains. Ordinarily, a polypeptide ECD will have less than about 1% of such transmembrane
15 and/or cytoplasmic domains and preferably, will have less than about 0.5% of such domains. It will be understood that any transmembrane domain(s) identified for the polypeptides of the present invention are identified pursuant to criteria routinely employed in the art for identifying that type of hydrophobic domain. The exact boundaries of a transmembrane domain may vary but most likely by no more than about 5 amino acids at either end of the
20 domain as initially identified and as shown in the appended figures. As such, in one embodiment of the present invention, the extracellular domain of a polypeptide of the present invention comprises amino acids 1 to X of the mature amino acid sequence, wherein X is any amino acid within 5 amino acids on either side of the extracellular domain/transmembrane domain boundary.

25 The approximate location of the “signal peptides” of the various PRO polypeptides disclosed herein are shown in the accompanying figures. It is noted, however, that the C-terminal boundary of a signal peptide may vary, but most likely by no more than about 5 amino acids on either side of the signal peptide C-terminal boundary as initially identified
30 herein, wherein the C-terminal boundary of the signal peptide may be identified pursuant to criteria routinely employed in the art for identifying that type of amino acid sequence element (*e.g.*, Nielsen *et al.*, Prot. Eng., 10:1-6 (1997) and von Heinje *et al.*, Nucl. Acids. Res., 14:4683-4690 (1986)). Moreover, it is also recognized that, in some cases, cleavage of a signal sequence from a secreted polypeptide is not entirely uniform, resulting in more than one

secreted species. These mature polypeptides, where the signal peptide is cleaved within no more than about 5 amino acids on either side of the C-terminal boundary of the signal peptide as identified herein, and the polynucleotides encoding them, are contemplated by the present invention.

5

A "polypeptide variant" of any one of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:4); ADAM8 (NM_001109) (SEQ ID NO:6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:10);
 10 NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:12); TM7SF1 (NM_003272) (SEQ ID NO:14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NO:16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:22); SLNAC1 (sodium channel receptor SLNAC1,
 15 NM_004769) (SEQ ID NO:24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:36);
 20 PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:42); or TCF4 (NM_030756) (SEQ ID NO:44) polypeptide as defined above or below having at least about 80% amino acid sequence identity with a full-length native sequence polypeptide, with or without the signal peptide, as
 25 disclosed herein or any other fragment of a full-length HGD marker polypeptides wherein one or more amino acid residues are added, or deleted, at the N- or C-terminus of the full-length native amino acid sequence. Ordinarily, a HGD marker polypeptide variant will have at least about 80% amino acid sequence identity, preferably at least about 81% amino acid sequence identity, more preferably at least about 82% amino acid sequence identity, more preferably at
 30 least about 83% amino acid sequence identity, more preferably at least about 84% amino acid sequence identity, more preferably at least about 85% amino acid sequence identity, more preferably at least about 86% amino acid sequence identity, more preferably at least about 87% amino acid sequence identity, more preferably at least about 88% amino acid sequence identity, more preferably at least about 89% amino acid sequence identity, more preferably at

least about 90% amino acid sequence identity, more preferably at least about 91% amino acid sequence identity, more preferably at least about 92% amino acid sequence identity, more preferably at least about 93% amino acid sequence identity, more preferably at least about 94% amino acid sequence identity, more preferably at least about 95% amino acid sequence identity, more preferably at least about 96% amino acid sequence identity, more preferably at least about 97% amino acid sequence identity, more preferably at least about 98% amino acid sequence identity and most preferably at least about 99% amino acid sequence identity with a full-length native sequence polypeptide sequence lacking the signal peptide as disclosed herein, an extracellular domain of a HGD marker polypeptide, with or without the signal peptide, as disclosed herein or any other fragment of a full-length HGD marker polypeptide sequence as disclosed herein. Ordinarily, a HGD marker polypeptide variant is at least about 10 amino acids in length, often at least about 20 amino acids in length, more often at least about 30 amino acids in length, more often at least about 40 amino acids in length, more often at least about 50 amino acids in length, more often at least about 60 amino acids in length, more often at least about 70 amino acids in length, more often at least about 80 amino acids in length, more often at least about 90 amino acids in length, more often at least about 100 amino acids in length, more often at least about 150 amino acids in length, more often at least about 200 amino acids in length, more often at least about 300 amino acids in length, or more.

"Percent (%) amino acid sequence identity" with respect to the amino acid sequence of any of the HGD marker polypeptides identified herein is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in an ET-1 (endothelin-1, NM_001955) (SEQ ID NO:2); AGR2 (anterior gradient 2 (Xenopus laevis) homolog, NM_006408) (SEQ ID NO:4); ADAM8 (NM_001109) (SEQ ID NO:6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:12); TM7SF1 (NM_003272) (SEQ ID NO:14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NO:16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID

NO:32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:42); or TCF4 (NM_030756) (SEQ ID NO:44) polypeptide, after
5 aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly
10 available computer software such as BLAST, BLAST-2, ALIGN, ALIGN-2 or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full-length of the sequences being compared. For purposes herein, however, % amino acid sequence identity values are obtained as described below by using the sequence comparison
15 computer program ALIGN-2, wherein the complete source code for the ALIGN-2 program is provided in Table 5. The ALIGN-2 sequence comparison computer program was authored by Genentech, Inc., and the source code shown in Table 5 has been filed with user documentation in the U.S. Copyright Office, Washington D.C., 20559, where it is registered under U.S. Copyright Registration No. TXU510087. The ALIGN-2 program is publicly available through
20 Genentech, Inc., South San Francisco, California or may be compiled from the source code provided in Table 5. The ALIGN-2 program should be compiled for use on a UNIX operating system, preferably digital UNIX V4.0D. All sequence comparison parameters are set by the ALIGN-2 program and do not vary.

25 For purposes herein, the % amino acid sequence identity of a given amino acid sequence A to, with, or against a given amino acid sequence B (which can alternatively be phrased as a given amino acid sequence A that has or comprises a certain % amino acid sequence identity to, with, or against a given amino acid sequence B) is calculated as follows:

30
$$100 \text{ times the fraction } X/Y$$

where X is the number of amino acid residues scored as identical matches by the sequence alignment program ALIGN-2 in that program's alignment of A and B, and where Y is the total number of amino acid residues in B. It will be appreciated that where the length of amino acid

sequence A is not equal to the length of amino acid sequence B, the % amino acid sequence identity of A to B will not equal the % amino acid sequence identity of B to A. As examples of % amino acid sequence identity calculations, Tables 2A-2B demonstrate how to calculate the % amino acid sequence identity of the amino acid sequence designated "Comparison Protein" to the amino acid sequence designated "PRO".

Unless specifically stated otherwise, all % amino acid sequence identity values used herein are obtained as described above using the ALIGN-2 sequence comparison computer program. However, % amino acid sequence identity may also be determined using the sequence comparison program NCBI-BLAST2 (Altschul *et al.*, Nucleic Acids Res., 25:3389-3402 (1997)). The NCBI-BLAST2 sequence comparison program may be downloaded from <http://www.ncbi.nlm.nih.gov>. NCBI-BLAST2 uses several search parameters, wherein all of those search parameters are set to default values including, for example, unmask = yes, strand = all, expected occurrences = 10, minimum low complexity length = 15/5, multi-pass e-value = 0.01, constant for multi-pass = 25, dropoff for final gapped alignment = 25 and scoring matrix = BLOSUM62.

In situations where NCBI-BLAST2 is employed for amino acid sequence comparisons, the % amino acid sequence identity of a given amino acid sequence A to, with, or against a given amino acid sequence B (which can alternatively be phrased as a given amino acid sequence A that has or comprises a certain % amino acid sequence identity to, with, or against a given amino acid sequence B) is calculated as follows:

$$100 \text{ times the fraction } X/Y$$

where X is the number of amino acid residues scored as identical matches by the sequence alignment program NCBI-BLAST2 in that program's alignment of A and B, and where Y is the total number of amino acid residues in B. It will be appreciated that where the length of amino acid sequence A is not equal to the length of amino acid sequence B, the % amino acid sequence identity of A to B will not equal the % amino acid sequence identity of B to A.

In addition, % amino acid sequence identity may also be determined using the WU-BLAST-2 computer program (Altschul *et al.*, Methods in Enzymology, 266:460-480 (1996)). Most of the WU-BLAST-2 search parameters are set to the default values. Those not set to

default values, *i.e.*, the adjustable parameters, are set with the following values: overlap span = 1, overlap fraction = 0.125, word threshold (T) = 11, and scoring matrix = BLOSUM62. For purposes herein, a % amino acid sequence identity value is determined by dividing (a) the number of matching identical amino acids residues between the amino acid sequence of the PRO polypeptide of interest having a sequence derived from the native PRO polypeptide and the comparison amino acid sequence of interest (*i.e.*, the sequence against which the PRO polypeptide of interest is being compared which may be a PRO variant polypeptide) as determined by WU-BLAST-2 by (b) the total number of amino acid residues of the PRO polypeptide of interest. For example, in the statement “a polypeptide comprising an amino acid sequence A which has or having at least 80% amino acid sequence identity to the amino acid sequence B”, the amino acid sequence A is the comparison amino acid sequence of interest and the amino acid sequence B is the amino acid sequence of the PRO polypeptide of interest.

As used herein, a “HGD marker” or “cancer marker gene or polypeptide,” or “anti-[HGD marker]” or “anti-[cancer marker]” refers to any one of the genes, polypeptides encoded by the genes, or antibodies specific for the polypeptides described herein as diagnostic for HGD or cancer. Thus, for example, “TCF4” refers to the gene marker or its encoded polypeptide, whereas anti-TCF4 refers to an antibody to the TCF4-encoded polypeptide.

A “gene variant polynucleotide” as used herein refers to a nucleic acid sequence that varies from the native sequence of its respective HGD marker gene NCBI accession sequence as disclosed in Table 4A, and further refers to a nucleic acid molecule which encodes a biologically active polypeptide and which nucleic acid molecule has at least about 80% nucleic acid sequence identity with a nucleic acid sequence selected from the group of marker genes: ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3); ADAM8 (NM_001109) (SEQ ID NO:5); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11); TM7SF1 (NM_003272) (SEQ ID NO:13); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NO:15); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21); SLNAC1 (sodium channel receptor SLNAC1, NM_004769)

(SEQ ID NO:23); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33);

5 CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41); and TCF4 (NM_030756) (SEQ ID NO:43), which genes encode, respectively, the full-length native polypeptides of the group:

10 ET-1 (endothelin-1, NM_001955) (SEQ ID NO:2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:4); ADAM8 (NM_001109) (SEQ ID NO:6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:12); TM7SF1 (NM_003272) (SEQ ID NO:14); DLDH (dihydrolipamide

15 dehydrogenase, NM_000108) (SEQ ID NO:16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:26); PA21 (phospholipase a2 precursor,

20 NM_000928) (SEQ ID NO:28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:38); CYB5 (cytochrome b5, 3' end,

25 NM_001914) (SEQ ID NO:40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:42); and TCF4 (NM_030756) (SEQ ID NO:44) polypeptide sequence as disclosed herein, a full-length native sequence HGD marker polypeptide sequence lacking the signal peptide as disclosed herein, an extracellular domain of a HGD marker polypeptide, with or without the signal peptide, as disclosed herein or any other fragment of a

30 full-length HGD marker polypeptide sequence as disclosed herein. Ordinarily, a HGD marker variant polynucleotide will have at least about 80% nucleic acid sequence identity, more preferably at least about 81% nucleic acid sequence identity, more preferably at least about 82% nucleic acid sequence identity, more preferably at least about 83% nucleic acid sequence identity, more preferably at least about 84% nucleic acid sequence identity, more preferably at

least about 85% nucleic acid sequence identity, more preferably at least about 86% nucleic acid sequence identity, more preferably at least about 87% nucleic acid sequence identity, more preferably at least about 88% nucleic acid sequence identity, more preferably at least about 89% nucleic acid sequence identity, more preferably at least about 90% nucleic acid sequence identity, more preferably at least about 91% nucleic acid sequence identity, more preferably at least about 92% nucleic acid sequence identity, more preferably at least about 93% nucleic acid sequence identity, more preferably at least about 94% nucleic acid sequence identity, more preferably at least about 95% nucleic acid sequence identity, more preferably at least about 96% nucleic acid sequence identity, more preferably at least about 97% nucleic acid sequence identity, more preferably at least about 98% nucleic acid sequence identity and yet more preferably at least about 99% nucleic acid sequence identity with the nucleic acid sequence encoding a full-length native sequence HGD marker polypeptide sequence as disclosed herein, a full-length native sequence HGD marker polypeptide sequence lacking the signal peptide as disclosed herein, an extracellular domain of a HGD marker polypeptide, with or without the signal sequence, as disclosed herein or any other fragment of a full-length HGD marker polypeptide sequence as disclosed herein. Variants do not encompass the native nucleotide sequence.

Ordinarily, HGD marker gene variant polynucleotides are at least about 20 nucleotides in length, frequently at least about 30 nucleotides in length, often at least about 60 nucleotides in length, more often at least about 90 nucleotides in length, more often at least about 120 nucleotides in length, more often at least about 150 nucleotides in length, more often at least about 180 nucleotides in length, more often at least about 210 nucleotides in length, more often at least about 240 nucleotides in length, more often at least about 270 nucleotides in length, more often at least about 300 nucleotides in length, more often at least about 450 nucleotides in length, more often at least about 600 nucleotides in length, more often at least about 900 nucleotides in length, or more.

"Percent (%) nucleic acid sequence identity" with respect to variant polypeptides of each of the HGD marker polypeptide-encoding nucleic acid sequences identified herein is defined as the percentage of nucleotides in a candidate sequence that are identical with the nucleotides in a HGD marker polypeptide-encoding nucleic acid sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent nucleic acid sequence identity can be

achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN, ALIGN-2 or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full-length of the sequences being compared. For purposes herein, however, % nucleic acid sequence identity values are obtained as described below by using the sequence comparison computer program ALIGN-2, wherein the complete source code for the ALIGN-2 program is provided in Table 5. The ALIGN-2 sequence comparison computer program was authored by Genentech, Inc., and the source code shown in Table 5 has been filed with user documentation in the U.S. Copyright Office, Washington D.C., 20559, where it is registered under U.S. Copyright Registration No. TXU510087. The ALIGN-2 program is publicly available through Genentech, Inc., South San Francisco, California or may be compiled from the source code provided in Table 5. The ALIGN-2 program should be compiled for use on a UNIX operating system, preferably digital UNIX V4.0D. All sequence comparison parameters are set by the ALIGN-2 program and do not vary.

For purposes herein, the % nucleic acid sequence identity of a given nucleic acid sequence C to, with, or against a given nucleic acid sequence D (which can alternatively be phrased as a given nucleic acid sequence C that has or comprises a certain % nucleic acid sequence identity to, with, or against a given nucleic acid sequence D) is calculated as follows:

$$100 \text{ times the fraction } W/Z$$

where W is the number of nucleotides scored as identical matches by the sequence alignment program ALIGN-2 in that program's alignment of C and D, and where Z is the total number of nucleotides in D. It will be appreciated that where the length of nucleic acid sequence C is not equal to the length of nucleic acid sequence D, the % nucleic acid sequence identity of C to D will not equal the % nucleic acid sequence identity of D to C. As examples of % nucleic acid sequence identity calculations, Tables 2C-2D demonstrate how to calculate the % nucleic acid sequence identity of the nucleic acid sequence designated "Comparison DNA" to the nucleic acid sequence designated "PRO-DNA".

Unless specifically stated otherwise, all % nucleic acid sequence identity values used herein are obtained as described above using the ALIGN-2 sequence comparison computer

program. However, % nucleic acid sequence identity may also be determined using the sequence comparison program NCBI-BLAST2 (Altschul *et al.*, Nucleic Acids Res., 25:3389-3402 (1997)). The NCBI-BLAST2 sequence comparison program may be downloaded from <http://www.ncbi.nlm.nih.gov>. NCBI-BLAST2 uses several search parameters, wherein all of those search parameters are set to default values including, for example, unmask = yes, strand = all, expected occurrences = 10, minimum low complexity length = 15/5, multi-pass e-value = 0.01, constant for multi-pass = 25, dropoff for final gapped alignment = 25 and scoring matrix = BLOSUM62.

In situations where NCBI-BLAST2 is employed for sequence comparisons, the % nucleic acid sequence identity of a given nucleic acid sequence C to, with, or against a given nucleic acid sequence D (which can alternatively be phrased as a given nucleic acid sequence C that has or comprises a certain % nucleic acid sequence identity to, with, or against a given nucleic acid sequence D) is calculated as follows:

$$100 \text{ times the fraction } W/Z$$

where W is the number of nucleotides scored as identical matches by the sequence alignment program NCBI-BLAST2 in that program's alignment of C and D, and where Z is the total number of nucleotides in D. It will be appreciated that where the length of nucleic acid sequence C is not equal to the length of nucleic acid sequence D, the % nucleic acid sequence identity of C to D will not equal the % nucleic acid sequence identity of D to C.

In addition, % nucleic acid sequence identity values may also be generated using the WU-BLAST-2 computer program (Altschul *et al.*, Methods in Enzymology, 266:460-480 (1996)). Most of the WU-BLAST-2 search parameters are set to the default values. Those not set to default values, *i.e.*, the adjustable parameters, are set with the following values: overlap span = 1, overlap fraction = 0.125, word threshold (T) = 11, and scoring matrix = BLOSUM62. For purposes herein, a % nucleic acid sequence identity value is determined by dividing (a) the number of matching identical nucleotides between the nucleic acid sequence of the PRO polypeptide-encoding nucleic acid molecule of interest having a sequence derived from the native sequence PRO polypeptide-encoding nucleic acid and the comparison nucleic acid molecule of interest (*i.e.*, the sequence against which the PRO polypeptide-encoding nucleic acid molecule of interest is being compared which may be a variant PRO polynucleotide) as determined by WU-BLAST-2 by (b) the total number of nucleotides of the

PRO polypeptide-encoding nucleic acid molecule of interest. For example, in the statement “an isolated nucleic acid molecule comprising a nucleic acid sequence A which has or having at least 80% nucleic acid sequence identity to the nucleic acid sequence B”, the nucleic acid sequence A is the comparison nucleic acid molecule of interest and the nucleic acid sequence B is the nucleic acid sequence of the PRO polypeptide-encoding nucleic acid molecule of interest.

In other embodiments, variants of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3); ADAM8 (NM_001109) (SEQ ID NO:5); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11); TM7SF1 (NM_003272) (SEQ ID NO:13); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41); or TCF4 (NM_030756) (SEQ ID NO:43) HGD marker genes encode an active HGD marker polypeptide, and nucleic acid sequences useful for identifying the marker genes by, for example, nucleic acid hybridization assays or PCR assays are capable of hybridizing, preferably under stringent hybridization and wash conditions, to nucleotide sequences encoding the full-length ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3); ADAM8 (NM_001109) (SEQ ID NO:5); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11); TM7SF1 (NM_003272) (SEQ ID NO:13); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15); MAT2B (methionine adenosyltransferase II, beta,

NM_013283) (SEQ ID NO:17); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41); and TCF4 (NM_030756) (SEQ ID NO:43) gene or hybridizable fragments thereof, which nucleotide sequences are found in the NCBI accession numbers listed in Table 4A for the respective polypeptides. HGD variant polypeptides may be those that are encoded by a HGD marker gene variant polynucleotide.

The term "positives", in the context of the amino acid sequence identity comparisons performed as described above, includes amino acid residues in the sequences compared that are not only identical, but also those that have similar properties. Amino acid residues that score a positive value to an amino acid residue of interest are those that are either identical to the amino acid residue of interest or are a preferred substitution (as defined in Table 4A below) of the amino acid residue of interest.

For purposes herein, the % value of positives of a given amino acid sequence A to, with, or against a given amino acid sequence B (which can alternatively be phrased as a given amino acid sequence A that has or comprises a certain % positives to, with, or against a given amino acid sequence B) is calculated as follows:

$$100 \text{ times the fraction } X/Y$$

where X is the number of amino acid residues scoring a positive value as defined above by the sequence alignment program ALIGN-2 in that program's alignment of A and B, and where Y is the total number of amino acid residues in B. It will be appreciated that where the length of amino acid sequence A is not equal to the length of amino acid sequence B, the % positives of A to B will not equal the % positives of B to A.

"Isolated," when used to describe the various polypeptides disclosed herein, means polypeptide that has been identified and separated and/or recovered from a component of its natural environment. Preferably, the isolated polypeptide is free of association with all components with which it is naturally associated. Contaminant components of its natural environment are materials that would typically interfere with diagnostic or therapeutic uses for the polypeptide, and may include enzymes, hormones, and other proteinaceous or non-proteinaceous solutes. In preferred embodiments, the polypeptide will be purified (1) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (2) to homogeneity by SDS-PAGE under non-reducing or reducing conditions using Coomassie blue or, preferably, silver stain. Isolated polypeptide includes polypeptide *in situ* within recombinant cells, since at least one component of the ET-1 (endothelin-1, NM_001955) (SEQ ID NO:2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:4); ADAM8 (NM_001109) (SEQ ID NO:6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:12); TM7SF1 (NM_003272) (SEQ ID NO:14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NO:16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:42); or TCF4 (NM_030756) (SEQ ID NO:44) polypeptide's natural environment will not be present. Ordinarily, however, isolated polypeptide will be prepared by at least one purification step.

An "isolated" nucleic acid molecule encoding an ET-1 (endothelin-1, NM_001955) (SEQ ID NO:2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID

NO:4); ADAM8 (NM_001109) (SEQ ID NO:6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:12); TM7SF1 (NM_003272) (SEQ ID NO:14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NO:16);

5 MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:28); PAR2

10 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:40); COXVIb (coxVIb gene, last exon

15 and flanking sequence, NM_001863) (SEQ ID NO:42); or TCF4 (NM_030756) (SEQ ID NO:44) polypeptide or an "isolated" nucleic acid encoding an anti-[HGD marker polypeptide] antibody, is a nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the HGD marker genes or the anti-[HGD marker polypeptide]-encoding nucleic acid.

20 Preferably, the isolated nucleic acid is free of association with all components with which it is naturally associated. An isolated polypeptide or nucleic acid sequence is other than in the form or setting in which it is found in nature. Isolated nucleic acid molecules therefore are distinguished from the nucleic acid molecule as it exists in natural cells. However, an isolated nucleic acid molecule encoding a HGD maker polypeptide or an anti-[HGD marker

25 polypeptide] antibody includes HGD marker gene nucleic acid molecules and anti-[HGD marker polypeptide]-encoding nucleic acid molecules contained in cells that ordinarily express HGD marker polypeptides or express anti-[HGD maker polypeptide] antibodies where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

30

The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence,

and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

The term "antibody" is used in the broadest sense and specifically covers, for example, single anti-[HGD marker polypeptide] monoclonal antibodies (including antagonist, and neutralizing antibodies), anti-[HGD marker polypeptide] antibody compositions with polypeptopic specificity, single chain anti-[HGD marker polypeptide] antibodies, and fragments thereof (see below). The term "monoclonal antibody" as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, *i.e.*, the individual antibodies comprising the population are identical except for possible naturally-occurring mutations that may be present in minor amounts.

"Stringency" of hybridization reactions is readily determinable by one of ordinary skill in the art, and generally is an empirical calculation dependent upon probe length, washing temperature, and salt concentration. In general, longer probes require higher temperatures for proper annealing, while shorter probes need lower temperatures. Hybridization generally depends on the ability of denatured DNA to reanneal when complementary strands are present in an environment below their melting temperature. The higher the degree of desired homology between the probe and hybridizable sequence, the higher the relative temperature which can be used. As a result, it follows that higher relative temperatures would tend to make the reaction conditions more stringent, while lower temperatures less so. For additional

details and explanation of stringency of hybridization reactions, *see* Ausubel *et al.*, Current Protocols in Molecular Biology, Wiley Interscience Publishers, (1995).

"Stringent conditions" or "high stringency conditions", as defined herein, may be identified by those that: (1) employ low ionic strength and high temperature for washing, for example 0.015 M sodium chloride/0.0015 M sodium citrate/0.1% sodium dodecyl sulfate at 50°C; (2) employ during hybridization a denaturing agent, such as formamide, for example, 50% (v/v) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 µg/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC (sodium chloride/sodium citrate) and 50% formamide at 55°C, followed by a high-stringency wash consisting of 0.1 x SSC containing EDTA at 55°C.

"Moderately stringent conditions" may be identified as described by Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, New York: Cold Spring Harbor Press, 1989, and include the use of washing solution and hybridization conditions (*e.g.*, temperature, ionic strength and % SDS) less stringent than those described above. An example of moderately stringent conditions is overnight incubation at 37°C in a solution comprising: 20% formamide, 5 x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5 x Denhardt's solution, 10% dextran sulfate, and 20 mg/ml denatured sheared salmon sperm DNA, followed by washing the filters in 1 x SSC at about 35°C-50°C. The skilled artisan will recognize how to adjust the temperature, ionic strength, etc. as necessary to accommodate factors such as probe length and the like.

The term "epitope tagged" when used herein refers to a chimeric polypeptide comprising a HGD marker polypeptide fused to a "tag polypeptide". The tag polypeptide has enough residues to provide an epitope against which an antibody can be made, yet is short enough such that it does not interfere with activity of the polypeptide to which it is fused. The tag polypeptide preferably also is fairly unique so that the antibody does not substantially cross-react with other epitopes. Suitable tag polypeptides generally have at least six amino

acid residues and usually between about 8 and 50 amino acid residues (preferably, between about 10 and 20 amino acid residues).

"Active" or "activity" for the purposes herein refers to form(s) of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:4); ADAM8 (NM_001109) (SEQ ID NO:6); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:12); TM7SF1 (NM_003272) (SEQ ID NO:14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NO:16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:18); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:42); or TCF4 (NM_030756) (SEQ ID NO:44) polypeptides which retain a biological and/or an immunological activity/property of a native or naturally-occurring HGD marker polypeptide, wherein "biological" activity refers to a function (either inhibitory or stimulatory) caused by a native or naturally-occurring HGD marker polypeptide other than the ability to induce the production of an antibody against an antigenic epitope possessed by a native or naturally-occurring HGD marker polypeptide and an "immunological" activity refers to the ability to induce the production of an antibody against an antigenic epitope possessed by a native or naturally-occurring HGD marker polypeptide.

"Biological activity" in the context of an antibody or another antagonist molecule, or therapeutic compound that can be identified by the screening assays disclosed herein (*e.g.*, an organic or inorganic small molecule, peptide, etc.) is used to refer to the ability of such molecules to bind or complex with the polypeptides encoded by the amplified genes identified herein, or otherwise interfere with the interaction of the encoded polypeptides with other

cellular proteins or otherwise interfere with the transcription or translation of a HGD marker polypeptide. "Biological activity" in the context of an agonist molecule that enhances the activity of, for example, native anti-angiogenic molecules refers to the ability of such molecules to bind or complex with the polypeptides encoded by the amplified genes identified
5 herein or otherwise modify the interaction of the encoded polypeptides with other cellular proteins or otherwise enhance the transcription or translation of a TIMP1 or thrombospondin 2 polypeptide. A preferred biological activity is growth inhibition of a target tumor cell. Another preferred biological activity is cytotoxic activity resulting in the death of the target tumor cell.

10

The term "biological activity" in the context of a HGD marker polypeptide means the typical activity of the HGD marker polypeptide in the cell.

15

The phrase "immunological activity" means immunological cross-reactivity with at least one epitope of a HGD marker polypeptide.

20

"Immunological cross-reactivity" as used herein means that the candidate polypeptide is capable of competitively inhibiting the qualitative biological activity of a HGD marker polypeptide having this activity with polyclonal antisera raised against the known active HGD
20 marker polypeptide. Such antisera are prepared in conventional fashion by injecting goats or rabbits, for example, subcutaneously with the known active analogue in complete Freund's adjuvant, followed by booster intraperitoneal or subcutaneous injection in incomplete Freund's. The immunological cross-reactivity preferably is "specific", which means that the binding affinity of the immunologically cross-reactive molecule (*e.g.*, antibody) identified, to the
25 corresponding HGD marker polypeptide is significantly higher (preferably at least about 2-times, more preferably at least about 4-times, even more preferably at least about 8-times, most preferably at least about 10-times higher) than the binding affinity of that molecule to any other known native polypeptide.

30

The term "antagonist" is used in the broadest sense, and includes any molecule that partially or fully blocks, inhibits, or neutralizes a biological activity of a native HGD marker polypeptide disclosed herein or the transcription or translation thereof, particularly when the HGD marker polypeptide is expressed about 1.5-fold above the level of expression in normal tissue controls. Suitable antagonist molecules specifically include antagonist antibodies or

antibody fragments, binding fragments, peptides, small organic molecules, anti-sense nucleic acids, etc. Included are methods for identifying antagonists of an ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1 or 2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3 or 4); ADAM8 (NM_001109) (SEQ ID NO:5 or 6); PRSS8

5 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7 or 8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9 or 10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11 or 12); TM7SF1 (NM_003272) (SEQ ID NO:13 or 14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15 or 16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17 or 18); STC-2

10 (stanniocalcin-2, NM_003714) (SEQ ID NO:19 or 20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21 or 22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23 or 24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25 or 26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27 or 28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID

15 NO:29 or 30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31 or 32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33 or 34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35 or 36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37 or 38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39 or 40); COXVIb (coxVIb gene, last exon and flanking

20 sequence, NM_001863) (SEQ ID NO:41 or 42); and TCF4 (NM_030756) (SEQ ID NO:43 or 44) gene or polypeptide with a candidate antagonist molecule and measuring a detectable change in one or more biological activities normally associated with the ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1 or 2); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3 or 4); ADAM8 (NM_001109) (SEQ ID NO:5 or 6); PRSS8

25 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7 or 8); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9 or 10); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11 or 12); TM7SF1 (NM_003272) (SEQ ID NO:13 or 14); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15 or 16); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17 or 18); STC-2

30 (stanniocalcin-2, NM_003714) (SEQ ID NO:19 or 20); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21 or 22); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23 or 24); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25 or 26); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27 or 28); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID

NO:29 or 30); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31 or 32); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33 or 34); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35 or 36); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37 or 38); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39 or 40); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41 or 42); and TCF4 (NM_030756) (SEQ ID NO:43 or 44) gene or polypeptide.

A "small molecule" is defined herein to have a molecular weight below about 500 Daltons.

"Antibodies" (Abs) and "immunoglobulins" (Igs) are glycoproteins having the same structural characteristics. While antibodies exhibit binding specificity to a specific antigen, immunoglobulins include both antibodies and other antibody-like molecules which lack antigen specificity. Polypeptides of the latter kind are, for example, produced at low levels by the lymph system and at increased levels by myelomas. The term "antibody" is used in the broadest sense and specifically covers, without limitation, intact monoclonal antibodies, polyclonal antibodies, multispecific antibodies (*e.g.*, bispecific antibodies) formed from at least two intact antibodies, and antibody fragments so long as they exhibit the desired biological activity.

"Native antibodies" and "native immunoglobulins" are usually heterotetrameric glycoproteins of about 150,000 daltons, composed of two identical light (L) chains and two identical heavy (H) chains. Each light chain is linked to a heavy chain by one covalent disulfide bond, while the number of disulfide linkages varies among the heavy chains of different immunoglobulin isotypes. Each heavy and light chain also has regularly spaced intrachain disulfide bridges. Each heavy chain has at one end a variable domain (V_H) followed by a number of constant domains. Each light chain has a variable domain at one end (V_L) and a constant domain at its other end; the constant domain of the light chain is aligned with the first constant domain of the heavy chain, and the light-chain variable domain is aligned with the variable domain of the heavy chain. Particular amino acid residues are believed to form an interface between the light- and heavy-chain variable domains.

The term "variable" refers to the fact that certain portions of the variable domains differ extensively in sequence among antibodies and are used in the binding and specificity of each particular antibody for its particular antigen. However, the variability is not evenly distributed throughout the variable domains of antibodies. It is concentrated in three segments
5 called complementarity-determining regions (CDRs) or hypervariable regions both in the light-chain and the heavy-chain variable domains. The more highly conserved portions of variable domains are called the framework (FR) regions. The variable domains of native heavy and light chains each comprise four FR regions, largely adopting a β -sheet configuration, connected by three CDRs, which form loops connecting, and in some cases
10 forming part of, the β -sheet structure. The CDRs in each chain are held together in close proximity by the FR regions and, with the CDRs from the other chain, contribute to the formation of the antigen-binding site of antibodies (*see* Kabat *et al.*, NIH Publ. No.91-3242, Vol. I, pages 647-669 (1991)). The constant domains are not involved directly in binding an antibody to an antigen, but exhibit various effector functions, such as participation of the
15 antibody in antibody-dependent cellular toxicity.

The term "hypervariable region" when used herein refers to the amino acid residues of an antibody which are responsible for antigen-binding. The hypervariable region comprises amino acid residues from a "complementarity determining region" or "CDR" (*i.e.*, residues
20 24-34 (L1), 50-56 (L2) and 89-97 (L3) in the light chain variable domain and 31-35 (H1), 50-65 (H2) and 95-102 (H3) in the heavy chain variable domain; Kabat *et al.*, Sequences of Proteins of Immunological Interest, 5th Ed. Public Health Service, National Institute of Health, Bethesda, MD. [1991]) and/or those residues from a "hypervariable loop" (*i.e.*, residues 26-32 (L1), 50-52 (L2) and 91-96 (L3) in the light chain variable domain and 26-32
25 (H1), 53-55 (H2) and 96-101 (H3) in the heavy chain variable domain ; Clothia and Lesk, J. Mol. Biol., 196:901-917 [1987]). "Framework" or "FR" residues are those variable domain residues other than the hypervariable region residues as herein defined.

"Antibody fragments" comprise a portion of an intact antibody, preferably the antigen
30 binding or variable region of the intact antibody. Examples of antibody fragments include Fab, Fab', F(ab')₂, and Fv fragments; diabodies; linear antibodies (Zapata *et al.*, Protein Eng., 8(10):1057-1062 [1995]); single-chain antibody molecules; and multispecific antibodies formed from antibody fragments.

Papain digestion of antibodies produces two identical antigen-binding fragments, called "Fab" fragments, each with a single antigen-binding site, and a residual "Fc" fragment, whose name reflects its ability to crystallize readily. Pepsin treatment yields an $F(ab')_2$ fragment that has two antigen-combining sites and is still capable of cross-linking antigen.

5

"Fv" is the minimum antibody fragment which contains a complete antigen-recognition and -binding site. This region consists of a dimer of one heavy- and one light-chain variable domain in tight, non-covalent association. It is in this configuration that the three CDRs of each variable domain interact to define an antigen-binding site on the surface of the V_H - V_L dimer. Collectively, the six CDRs confer antigen-binding specificity to the antibody. However, even a single variable domain (or half of an Fv comprising only three CDRs specific for an antigen) has the ability to recognize and bind antigen, although at a lower affinity than the entire binding site.

The Fab fragment also contains the constant domain of the light chain and the first constant domain (CH1) of the heavy chain. Fab fragments differ from Fab' fragments by the addition of a few residues at the carboxy terminus of the heavy chain CH1 domain including one or more cysteines from the antibody hinge region. Fab'-SH is the designation herein for Fab' in which the cysteine residue(s) of the constant domains bear a free thiol group. $F(ab')_2$ antibody fragments originally were produced as pairs of Fab' fragments which have hinge cysteines between them. Other chemical couplings of antibody fragments are also known.

The "light chains" of antibodies (immunoglobulins) from any vertebrate species can be assigned to one of two clearly distinct types, called kappa (κ) and lambda (λ), based on the amino acid sequences of their constant domains.

Depending on the amino acid sequence of the constant domain of their heavy chains, immunoglobulins can be assigned to different classes. There are five major classes of immunoglobulins: IgA, IgD, IgE, IgG, and IgM, and several of these may be further divided into subclasses (isotypes), *e.g.*, IgG1, IgG2, IgG3, IgG4, IgA1, and IgA2. The heavy-chain constant domains that correspond to the different classes of immunoglobulins are called α , δ , ϵ , γ , and μ , respectively. The subunit structures and three-dimensional configurations of different classes of immunoglobulins are well known.

The term "monoclonal antibody" as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, *i.e.*, the individual antibodies comprising the population are identical except for possible naturally occurring mutations that may be present in minor amounts. Monoclonal antibodies are highly specific, being directed against a single antigenic site. Furthermore, in contrast to conventional (polyclonal) antibody preparations which typically include different antibodies directed against different determinants (epitopes), each monoclonal antibody is directed against a single determinant on the antigen. In addition to their specificity, the monoclonal antibodies are advantageous in that they are synthesized by the hybridoma culture, uncontaminated by other immunoglobulins. The modifier "monoclonal" indicates the character of the antibody as being obtained from a substantially homogeneous population of antibodies, and is not to be construed as requiring production of the antibody by any particular method. For example, the monoclonal antibodies to be used in accordance with the present invention may be made by the hybridoma method first described by Kohler *et al.*, Nature, 256:495 [1975], or may be made by recombinant DNA methods (*see, e.g.*, U.S. Patent No. 4,816,567). The "monoclonal antibodies" may also be isolated from phage antibody libraries using the techniques described in Clackson *et al.*, Nature, 352:624-628 [1991] and Marks *et al.*, J. Mol. Biol., 222:581-597 (1991), for example.

The monoclonal antibodies herein specifically include "chimeric" antibodies (immunoglobulins) in which a portion of the heavy and/or light chain is identical with or homologous to corresponding sequences in antibodies derived from a particular species or belonging to a particular antibody class or subclass, while the remainder of the chain(s) is identical with or homologous to corresponding sequences in antibodies derived from another species or belonging to another antibody class or subclass, as well as fragments of such antibodies, so long as they exhibit the desired biological activity (U.S. Patent No. 4,816,567; Morrison *et al.*, Proc. Natl. Acad. Sci. USA, 81:6851-6855 [1984]).

"Humanized" forms of non-human (*e.g.*, murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')₂ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. For the most part, humanized antibodies are human immunoglobulins (recipient antibody) in which residues from a CDR of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat

or rabbit having the desired specificity, affinity, and capacity. In some instances, Fv FR residues of the human immunoglobulin are replaced by corresponding non-human residues. Furthermore, humanized antibodies may comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. These modifications are made to further refine and maximize antibody performance. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin. For further details, *see*, Jones *et al.*, Nature, 321:522-525 (1986); Reichmann *et al.*, Nature, 332:323-329 [1988]; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992). The humanized antibody includes a PRIMATIZEDTM antibody wherein the antigen-binding region of the antibody is derived from an antibody produced by immunizing macaque monkeys with the antigen of interest.

"Single-chain Fv" or "sFv" antibody fragments comprise the V_H and V_L domains of antibody, wherein these domains are present in a single polypeptide chain. Preferably, the Fv polypeptide further comprises a polypeptide linker between the V_H and V_L domains which enables the sFv to form the desired structure for antigen binding. For a review of sFv *see* Pluckthun in The Pharmacology of Monoclonal Antibodies, vol. 113, Rosenberg and Moore eds., Springer-Verlag, New York, pp. 269-315 (1994).

The term "diabodies" refers to small antibody fragments with two antigen-binding sites, which fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) in the same polypeptide chain (V_H - V_L). By using a linker that is too short to allow pairing between the two domains on the same chain, the domains are forced to pair with the complementary domains of another chain and create two antigen-binding sites. Diabodies are described more fully in, for example, EP 404,097; WO 93/11161; and Hollinger *et al.*, Proc. Natl. Acad. Sci. USA, 90:6444-6448 (1993).

An "isolated" antibody is one which has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials which would interfere with diagnostic or therapeutic uses for the

antibody, and may include enzymes, hormones, and other proteinaceous or nonproteinaceous solutes. In preferred embodiments, the antibody will be purified (1) to greater than 95% by weight of antibody as determined by the Lowry method, and most preferably more than 99% by weight, (2) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (3) to homogeneity by SDS-PAGE under reducing or nonreducing conditions using Coomassie blue or, preferably, silver stain. Isolated antibody includes the antibody *in situ* within recombinant cells since at least one component of the antibody's natural environment will not be present. Ordinarily, however, isolated antibody will be prepared by at least one purification step.

The word "label" when used herein refers to a detectable compound or composition which is conjugated directly or indirectly to the antibody so as to generate a "labeled" antibody. The label may be detectable by itself (*e.g.*, radioisotope labels or fluorescent labels) or, in the case of an enzymatic label, may catalyze chemical alteration of a substrate compound or composition which is detectable. Radionuclides that can serve as detectable labels include, for example, I-131, I-123, I-125, Y-90, Re-188, Re-186, At-211, Cu-67, Bi-212, and Pd-109. The label may also be a non-detectable entity such as a toxin.

A "liposome" is a small vesicle composed of various types of lipids, phospholipids and/or surfactant which is useful for delivery of a drug (such as a CXCR4; Laminin alpha 4; TIMP1; Type IV collagen alpha 1; Laminin alpha 3; Adrenomedullin; Thrombospondin 2; Type I collagen alpha 2; Type VI collagen alpha 2; Type VI collagen alpha 3; Latent TGFbeta binding protein 2 (LTBP2); Serine or cysteine protease inhibitor heat shock protein (HSP47); Procollagen-lysine, 2-oxoglutarate 5-dioxygenase; connexin 43; Type IV collagen alpha 2; Connexin 37; Ephrin A1; Laminin beta 2; Integrin alpha 1; Stanniocalcin 1; Thrombospondin 4; or CD36 polypeptide or antibody thereto and, optionally, a chemotherapeutic agent) to a mammal. The components of the liposome are commonly arranged in a bilayer formation, similar to the lipid arrangement of biological membranes.

As used herein, the term "immunoadhesin" designates antibody-like molecules which combine the binding specificity of a heterologous protein (an "adhesin") with the effector functions of immunoglobulin constant domains. Structurally, the immunoadhesins comprise a fusion of an amino acid sequence with the desired binding specificity which is other than the antigen recognition and binding site of an antibody (*i.e.*, is "heterologous"), and an

immunoglobulin constant domain sequence. The adhesin part of an immunoadhesin molecule typically is a contiguous amino acid sequence comprising at least the binding site of a receptor or a ligand. The immunoglobulin constant domain sequence in the immunoadhesin may be obtained from any immunoglobulin, such as IgG-1, IgG-2, IgG-3, or IgG-4 subtypes, IgA (including IgA-1 and IgA-2), IgE, IgD or IgM.

"Up-regulation," "increased expression," and "overexpression" are used interchangeably and, as used herein, mean at least about a 1.5-fold increase in expression, alternatively at least about a 2-fold increase in expression, alternatively with at least about a 2.5-fold or higher increase in expression of a gene measured as an increase in its DNA (amplification), its mRNA (increased transcription), or in the level of polypeptide encoded by the gene. Alternatively, up-regulation or increased expression is determined using a Z score as a p value < 0.07 relative to a normal tissue control.

The term "package insert" is used to refer to instructions customarily included in commercial packages of therapeutic products, that contain information about the indications, usage, dosage, administration, contraindications and/or warnings concerning the use of such therapeutic products.

It will be clearly understood that, although a number of art publications are referred to herein, this reference does not constitute an admission that any of these documents forms part of the common general knowledge in the art, in Australia or in any other country.

Throughout this specification and the claims, the terms "comprise," "comprises," and "comprising" are used in a non-exclusive sense, except where the context requires otherwise.

EXAMPLES

The following examples are offered by way of illustration and not by way of limitations. The examples are provided so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the compounds, compositions, and methods of the invention and are not intended to limit the scope of what the inventors regard as their invention. Efforts have been made to insure accuracy with respect to numbers used (e.g. amounts, temperature, etc. but some experimental errors and deviation should be

accounted for. Unless indicated otherwise, parts are in parts by weight, temperature is in degrees C, and pressure is at or near atmospheric. The disclosures of all citations in the specification are expressly incorporated herein by reference.

5 **Example 1: Patients and Tissue Collection**

Esophageal mucosal biopsies were obtained from patients undergoing surveillance endoscopy at the Western General Hospital and Royal Infirmary, Edinburgh during 2000-1. The study was approved by the Lothian Research and Ethics Committee and written, informed consent was obtained from all patients. All procedures were performed by one of two
10 experienced endoscopists with expertise in Barrett's esophagus in a standard manner according to a local protocol for Barrett's surveillance. BE was defined as tongues or circumferential salmon pink mucosa extending for at least 3cm above the gastro-esophageal junction. At endoscopy, careful note was made of the length of the CE segment, severity of any esophagitis if present and the presence of macroscopically visible abnormalities within the BE. Data on
15 smoking history, use of acid-suppressing drugs and *Helicobacter pylori* status were also recorded.

Paired biopsies were taken. One sample was fixed in formalin for histology and the other stored fresh-frozen (-70°C) for microarray analysis. Two gastrointestinal pathologists
20 reviewed all specimens, which were categorized as: normal squamous esophagus, BE (columnar lined esophagus with intestinal metaplasia and the presence of goblet cells and alcian blue positive mucin), BE with changes indeterminate dysplasia, BE with low-grade dysplasia (LGD), BE with high-grade dysplasia (HGD) or BE with adenocarcinoma (CA). For some patients, 2 separate biopsy specimens for the same disease state were available for array
25 analysis. Additional matched samples were also analyzed (e.g. biopsies of BE adjacent to carcinoma in BE from the same patient). Analyzed samples included 10 normal esophagus, 28 samples of BE from 20 patients, 6 samples of LGD from 3 patients, 3 samples indeterminate for dysplasia from 2 patients, 6 samples HGD from 3 patients, 10 samples of BE adjacent to CA (BE-CA) from 7 patients, 16 samples CA from 10 patients.

30

Microarrays containing 9031 genes were generated by printing PCR products derived from cDNA clones (Invitrogen, California and Genentech, Inc.) on glass slides coated with 3-aminopropyltriethoxysilane (Aldrich, Milwaukee WI) and 1,4-phenylenediisothiocyanate (Aldrich, Milwaukee WI) using a robotic arrayer (Norgren Systems, Mountain View,

California). RNA isolation was accomplished by CsCl step gradient, (Kingston, Current Protocols in Molecular Biology 1:4.2.5-4.2.6 (1998)) typically 0.1 – 2 µg of total RNA was obtained. Probes for array analysis were generated by conservative amplification and subsequent labelling as follows: double-stranded DNA generated from 0.1 µg of total RNA (Invitrogen, Carlsbad, CA) was amplified using a single round of a modified in vitro transcription protocol (MEGAScript T7 from Ambion, Austin, Texas (Gelder et al., Proc. Natl. Acad. Sci. USA 87:1663-1667 (1990))). The resulting cRNA was used as a template to generate a sense DNA probe using random primers (9mers, 0.15 mg/ml), Alexa 488 dUTP or Alexa 546 dUTP (40 µM and 6 µM, respectively, Molecular Probes, Eugene, Oregon) using MMLV-derived reverse transcriptase (Invitrogen, Carlsbad, CA). A reference probe to reflect general epithelial cell expression was generated from 0.1 µg of total RNA from a pool of liver, lung and kidney (Clontech, Palo Alto, California). Probes were hybridized to arrays overnight in 50% formamide / 5XSSC at 37 °C and washed the next day in 2XSSC, 0.2% SDS followed by 0.2XSSC, 0.2% SDS. Array images were collected using a CCD-camera based imaging system (Norgren Systems, Mountain View, California) equipped with a Xenon light source and optical filters appropriate for each dye. Full dynamic-range images were collected (Autograb, Genentech Inc) and intensities and ratios extracted using automated gridding and data extraction software (gImage, Genentech Inc) built on a Matlab (the MathWorks, Natick, Massachusetts) platform.

Example 3: Data Analysis

Data were sorted to identify genes expressed above background (N intensity of > 12 where background values range from 0 – 8) in the test sample such that only meaningful ratios were included. Ratio values were further normalized for experimental scatter at different intensity values within each experiment by plotting log ratio versus N intensity and by fitting a normal distribution at each intensity level. A measure of standard deviation (Z score) around a mean of zero was derived for each gene in each experiment and this value was used in data mining. Specifically, for each microarray, data were normalized by computing Z-scores, which were obtained from a scatterplot of the logarithm of the ratio of the test and reference data versus the logarithm of the minimum of the test and reference data. The median of the ratio as a function of intensity was estimated by applying the loess algorithm to the scatterplot. The standard error was estimated by applying loess to the square root of the absolute residuals, and squaring the result to obtain the median absolute deviation (MAD), and making a

multiplicative correction to convert from MAD to a standard error. The Z scores were determined for each ratio by dividing its vertical distance from the median loess curve by the standard error at that intensity.

5 A computational process useful computing Z-scores may be written in a standard high-level statistical language, S-Plus, as follows:

```

pos.test <- test[test > 0 & ref > 0]
pos.ref <- ref[test > 0 & ref > 0]
10 minorder <- order(pmin(pos.test,pos.ref))
y <- log(pos.test[minorder] + 10) - log(pos.ref[minorder] + 10)
x <- log(pmin(pos.test[minorder],pos.ref[minorder]))
residuals <- loess(y ~ x)$residuals
sqresiduals <- sqrt(abs(residuals))
15 sqrt.mad <- loess(sqresiduals ~ x)$fitted
sigma <- sqrt.mad*sqrt.mad/0.6745
zscore <- ifelse(sigma > 0,residuals/sigma,0)

```

This code may be executed in a commercially available S-Plus program such as, for example,
20 (<http://www.insightful.com>), or in a freely available substitute program, R (<http://www.r-project.org>).

Example 4: Differential Expression in Barrett's Esophagus-to-Adenocarcinoma Disease Stages

25

Samples and Data Mining:

High-quality data were obtained from > 90% of biopsy specimens, including those of poor RNA quality and very limited RNA quantity (eg. less than 200 ng total RNA). A data
30 mining strategy was applied to identify genes specifically associated with the different stages of disease progression. Experiments were grouped into disease categories based on pathologic diagnosis, and these groups compared to identify genes with significant elevated expression for at least 25% of the samples within a disease group with respect to both the epithelial pool reference and the normal esophagus group. Typically, genes with elevated expression were

identified as those with Z scores of > 1.7 ($p < 0.05$) in the disease group, corresponding to ratio values of 2 – 20 in most cases. A total of 460 genes satisfied these criteria across the disease groups BE, dysplasia, and carcinoma (some genes are associated with more than one disease group). Selected genes (117) are listed (Tables 1, 2, 3). All dysplasia samples (high-, low-grade and indeterminate) were combined into a single group to improve data analysis, and the genes identified were then further inspected to determine if they were more prevalent in low- or high-grade dysplasia. HGD sample data were independently analyzed to determine gene expression profiles diagnostic for high-grade dysplasia (Table 4A).

Inflammation:

Significant expression of proinflammatory, costimulatory and inducible cytokines and receptors was observed in BE, dysplasia and carcinoma, and the most prevalent genes are listed (Table 1). Some binding partners were detected, such as putative inflammatory cytokine IL-17 family member IL-17E and its receptor IL-17BR, and SCYA20/LARC and receptor CCR6 (Lee et al., J. Biol. Chem. 276:1660-1664 (2001); and Baba et al., J. Biol. Chem. 272:14893-14898 (1997)). SCYA20 is expressed in the epithelium of the small intestine and is chemotactic for lymphocytes and dendritic cells (Tanaka et al., Eur. J. Immunol. 29:644-642 (1999)). Activin A is a TGF beta superfamily member that can act as a potent mediator of cell growth and differentiation and may be involved in response to injury (Munz et al., EMBO J. 18:5205-5215 (1999)). It was co-expressed particularly in carcinoma in Barrett's samples with its serine-threonine kinase receptor AVRII (the type I receptor was also detected but less well correlated). Chemokine receptors CXCR4 and CCR7 have been detected on a variety of inflammatory cell types, but have also been described as highly expressed in breast tumor cells, with possible involvement in lymph node metastasis (Muller et al., Nature 410:50-56 (2001)). In this study, CXCR4 in particular was associated with high-grade dysplasia and detected in some samples of adenocarcinoma.

TABLE 1A Cytokines and chemokines up-regulated in BE-to-Adenocarcinoma

NCBI RefSeq	Gene	BE	D	BE-CA	CA
NM_000594	TNF-a	*		*	*
NM_002546	Osteoprotegerin	*		*	
NM_002993	GCP-2	(*)	* H	(*)	*
NM_025240	B7-H3		* L	(*)	*
NM_002995	Lymphotactin	(*)	*		(*)
NM_005746	PBEF	*			(*)
NM_004591	SCYA20		(*)	*	
NM_004843	WSX1		*		
NM_019618	IL1-H1	(*)		*	*
NM_000418	IL-4R				*
NM_022789	IL-17E	(*)	*	*	*
NM_018725	IL-17BR		* H		(*)
NM_014432	IL-20Ra		* L		(*)
NM_021798	IL-21R	(*)		*	*
NM_002192	Activin A		(*)	(*)	*
NM_001616	AVR2, type II activin receptor		*		*
NM_001105	Activin A type I Receptor				(*)
NM_031409	CCR6	(*)		*	*
NM_003467	CXCR4		* H		(*)
NM_001838	CKR7	(*)	(*)	*	

TABLE 1B Prostaglandin synthesis-related genes up-regulated in BE-to-Adenocarcinoma

NCBI RefSeq	Gene	BE	D	BE-CA	CA
NM_000963	COX-2, prostaglandin synthase 2	(*)	* H		*
NM_000962	COX-1, prostaglandin synthase 1				*
NM_007366	PLA2R phospholipase A2 R1		*	(*)	*
NM_000953	PD2R prostaglandin D2 R	(*)		(*)	*
NM_000959	PF2AR prostaglandin F2 α R		*	(*)	(*)
NM_000957	PER3 prostaglandin E R 2			(*)	*
NM_000960	Prostaglandin IP (I2) R	*	*	(*)	

Genes are associated with the disease states B3, dysplasia (D), BE adjacent to carcinoma (BE-CA), or carcinoma (CA) if present in at least 25% of samples tested. (*) indicates gene expression changes associated with 15-25% of samples.

5 An otherwise rare IL-1 homolog, IL1-H1, was highly expressed in carcinoma in Barrett's, and also the matched adjacent BE tissue from the same patients (Fig. 1). A previous study of the murine IL-1H1 ortholog detected constitutive only in esophageal squamous mucosa. In addition, human IL1-H1 mRNA could be induced in TNF α and IFN α treated keratinocytes and squamous epithelial tumor cell line A431 (Kumar et al., J. Biol. Chem. 10 275:10308-10314 (2000)). This gene is one marker of a specific esophageal squamous cell type exhibiting a striking induction of expression in both adenocarcinoma and patient-matched BE, amidst primarily intestinal and tumor markers observed in this study (Tables 2 and 3). The high expression in BE matched with adenocarcinoma in addition to adenocarcinoma suggests a possible epigenetic association.

15 Cyclooxygenase isoform 2 (COX-2), which catalyzes a rate-limiting step in conversion of arachidonate to inflammatory prostaglandins, has been implicated in Barrett's metaplasia and other cancers (Morris et al., Am. J. Gastroenterol. 96:990-996 (2001); Heasley et al., J. Biol. Chem. 272:14501-14504 (1997); and Tsujii et al., Cell 93:705-716 (1998)). Consistent 20 with previous reports, a significant increase was observed in COX-2 gene expression with increasing dysplasia (high-grade dysplasia) and in adenocarcinoma (Table 1B). Smaller changes were also observed in COX-1 and several prostaglandin receptors. Arachidonic acid is released from the membrane by the action of phospholipases. Phospholipase A2 expression associated with increasing malignancy was also observed (Table 2) along with the M-type 25 receptor (PLA2R, Table 1B), consistent with studies suggesting that COX-2, PA2 and PLA2R are coordinately expressed (Rys-Sikora et al., Am. Physiol. Cell Physiol. 278:822-833 (2000)).

 Elevated expression was detected for another enzyme that generates a different class of biologically active eicosanoids from arachidonic acid, the epoxygenase CYP2J2 (Fig. 1B, 30 Table 2). This cytochrome P450 enzyme is expressed in a variety of cell types in the small intestine, including epithelial cells, and may play a role in electrolyte transport, intestinal motility, and other processes (Wu et al., J. Biol. Chem. 271:3460-3468 (1996); Zeldin et al., Mol. Pharm. 51:931-943 (1997); and Node et al., Science 285:1276-1279 (1999)). Similar to COX-2, elevated expression is most apparent in samples of adenocarcinoma and dysplasia

(both low-grade and high-grade dysplasia). The expression profile for CYP2J2 also reflects the progressive intestinal metaplasia observed in this study (Table 2).

Intestinal Metaplasia:

5

Analysis for gene expression changes associated with dysplasia revealed a large group of genes whose normal expression is primarily associated with the small intestine, and to a lesser extent, colon (Table 2). The previously described marker villin was detected, (Peterson and Moosekar, J. Cell Sci. 102:581-600 (1992)) along with a diverse set of genes including
10 cell surface cadherins and claudins, ion channels and transporters, and enzymes, many of which are normally associated with structural and absorptive functions of small intestinal villi. Increased expression of many of these genes was associated with dysplasia and a significant subset of carcinoma samples, with differential expression also detected in a smaller subset of BE samples. Furthermore, expression of the majority of genes was less prevalent in matched
15 BE samples taken from the carcinoma patients, even when expression was apparent in the tumor sample (Fig. 2A, 2B, 3A; Table 2). This suggests that these gene expression changes are more specifically associated with the foci of dysplasia and developing carcinoma within the larger region of BE.

TABLE 2 Genes up-regulated in intestinal metaplasia

NCBI RefSeq	SEQ ID NOS (na and aa)	Gene	Gene Description	BE	D	BE-CA	CA	Normal Tissues
NM_007127		Villin 1	actin binding protein	*	*	*	*	SI, C
NM_003379		Villin 2	actin binding protein	*				SI, St, C, O
NM_000775	35 and 36	CYP2J2	arachidonic acid epoxigenase		*	(*)	*	SI, L, H
NM_005379	33 and 34	MYO1A	myosin 1A		* H		*	SI (C)
NM_004063	45 and 46	CAD17	liver-intestine cadherin	(*)	(* H)	(*)	*	SI, C
NM_017717		MUCDHL	mucin and cadherin like			*		SI (C, K)
NM_014343	47 and 48	CLDN15	claudin 15	(*)	* L	(*)	*	SI
NM_012132		CLDN8	claudin 8		*		(*)	C, K
NM_005567		IR-95	lectin-binding			(*)	*	C, SI, St, O
NM_000021		Presenilin-1	beta-catenin binding		* H		(*)	SI, C
NM_003039		GLUT5	glucose transporter	*	(*)		(*)	SI
NM_001081		CUBN	transport (HDL, vit.B12, etc)		* L			K, SI
NM_004769	23 and 24	SLNAC1	sodium channel		* H	*	*	CNS, SI, O
NM_000492	49 and 50	CFTR	chloride channel	*	(* H)		*	P, SI, C
NM_003272	13 and 14	TM7SF1	novel GPCR	(*)	* H			K, C, SI, O
NM_005242	29 and 30	PAR2 / F2RL1	GPCR, proteinase-activated		* H			SI, C
NM_022304	51 and 52	H2R	histamine H2 receptor	(*)	*	*	*	St-par
NM_004624		VIPR1	intestinal peptide GPCR			*		L, SI, C, CNS

NM_002773	7 and 8	PRSS8	serine protease			*	* SI, C, St
NM_058186		RPLA320	novel		* L	(*)	SI (St, C, P)
NM_003561		SPLA2	phospholipase A2 group X		*	(*)	C, St, SI
NM_000928	27 and 28	PA21	phospholipase A2 group IB		*	(*)	P, SI, C
NM_001631	21 and 22	PPBI	intestinal alkaline phosphatase	(*)	*		SI
NM_000717	25 and 26	CAH4	carbonic anhydrase IV		* H	(*)	C, SI
NM_005763		LKR/SDH	lysine catabolism	(*)	* H		SI, C, O
NM_004969	31 and 32	IDE	insulin degrading enzyme	(*)	*	*	SI-ent., O
NM_001914	39 and 40	CYB5	cytochrome B5	(*)	* H		L, SI, K
NM_001863	41 and 42	COX6B	cytochrome C oxidase subunit	(*)	* H	*	H, M, SI, C, St
NM_000108	15 and 16	DLDH	dihydrolipamide dehydrogenase	(*)	*		H, M, K; SI, C
NM_006214	37 and 38	PHYH	phytanoyl-CoA hydroxylase		* H		L, K, M; SI, C
NM_013283	17 and 18	MAT2B	methionine adenosyltransferase		* H	(*)	SI, C, O
NM_000414		BHSD	hydroxysteroid dehydrogenase			(*)	L, SI, O
NM_005038		cyclophilin-40	peptidyl prolyl isomerase		* L	*	SI, C, L, M
NM_138393		DP1	membrane trafficking		(*)	*	L, SI
NM_006408	3 and 4	AGR2	anterior gradient 2 homolog		* H	*	St, SI, C
NM_021969	11 and 12	NROB2	nuclear hormone receptor	*	* H	*	SI, L, St
NM_005524		Hes1	transcriptional regulator	*	* H	*	SI-ent., O
NM_002054		GCG	proglucagon		(*)	*	P, SI, C

Genes are associated with the disease states B3, dysplasia (D), BE adjacent to carcinoma (BE-CA), or carcinoma (CA) if present in at least 25% of samples tested. (*) indicates gene expression changes associated with 15-25% of samples.

5

Normal Tissues: highest normal tissue expression is listed. SI (small intestine); C (colon); St (stomach); K (kidney); P (pancreas); L (liver); M (muscle); H (heart); CNS (central nervous system); SI-ent (intestinal enterocytes); St-par (parietal cells); O (other tissues). In the dysplasia column, H or L denote expression associated with high-grade or low-grade dysplasia, respectively. GPCR (G protein coupled receptor). "na" and "aa" refer to the nucleic acid and amino acid SEQ ID NO, respectively, for the associated markers.

Examples include MYO1A, an unconventional myosin that is differentially expressed along with crypt-villus axis, exhibiting low level cytosolic expression in immature crypts and high expression in villus cells with localization at the brush border (Skowron et al., Cell Motil Cytoskel. 41:308-324 (1998); and MacLennan et al., Molec. Carcinogen. 24:137-143 (1999)). Unlike villin, another marker of the brush border that was detected across all disease states, MYO1A was most associated with high-grade dysplasia and carcinoma. The novel secreted factor AGR2 gives one of the most striking profiles as a marker for high-grade dysplasia (Figure 2A). AGR2 is a human homolog of the *X. laevis* cement gland gene XAG-2, which is implicated in ectodermal patterning (Aberger et al., Mech. Dev. 72:115-130 (1998)). Elevated expression of this gene is also associated with hormonally-responsive high-grade esophageal dysplasias (Thompson and Weigel, Biochem. Biophys. Res. Commun. 251:111-116 (1998)).

Expression of nuclear hormone receptor NROB2 is induced by bile acids, and NROB2 in turn participates in transcriptional repression of the rate-limiting enzyme (CYP7A1) in bile synthesis (Lu et al., Mol. Cell 6:507-515 (2000)). In this study, overexpression of NROB2 is detected in particular in high-grade dysplasia, in addition to some carcinomas and a subset of BE samples (Figure 2B). In addition to supporting the general pattern of intestinal metaplasia, expression of NROB2 may further reflect the response to the unnatural exposure of esophageal cells to bile, which is considered to be a contributing factor in Barrett's metaplasia (Bremner et al, Surgery 68:209-216 (1970); and Gillen et al., Br. J. Surg. 75:1352-1355 (1988)). Bile acids have also been shown to activate transcription of COX-2 (Zhang et al., J. Biol. Chem. 273:2424-2428 (1998)).

While these gene expression profiles are consistent with the observations of an increased columnar cell type in BE, the most consistent changes are associated with dysplasia, especially high-grade dysplasia (Table 2). These genes could serve as markers for progression in a clinical setting. For example, the number of genes which meet the described criteria for elevated expression in individual samples progressively increases through BE and dysplasia. The average of the number of markers detected per sample is 7.6 for BE, 11.7 for low-grade dysplasia, and 16.4 for high-grade dysplasia. Within the BE group, 3 samples have unusually high scores of 12, 12, and 14 markers detected. The two samples with 12 markers are different biopsies from the same patient: while the overall expression profiles vary between the 2 biopsies, they score identically in the marker analysis. Marker selection could be further refined to a subset associated with particular disease stages. This type of quantitative analysis may be of utility in identifying BE patients with greater risk of progression, and may be less sensitive to sampling and observer-related effects. Some of the secreted and processed factors listed (Table 1A, 2, 3) may even be detectable in the blood, which could further simplify screening.

Adenocarcinoma:

Many of the genes differentially expressed in adenocarcinoma in Barrett's, similar to other solid tumors, reflect the changes occurring as the cells acquire a more proliferative and invasive phenotype (Table 3). Included are genes involved with growth, cell adhesion, matrix invasion, vascularization, and intracellular remodeling. The majority of genes are most prevalent in adenocarcinoma, but some are also detected at earlier stages. For example, genes likely to be involved in tumor angiogenesis showed significant upregulation in samples with dysplasia (eg. tumor endothelial marker 1 (TEM1), Tie2 ligand 2, VEGFC, endothelin 1).

TABLE 3 Genes up-regulated in esophageal adenocarcinoma

NCBI RefSeq	Gene families/genes	BE	D	BE-CA	CA
Growth factors / receptors					
NM_005228	EGFR		(* H)		*
NM_004442	EPHB2				*
NM_003212	CRIPTO CR-1	(*)	*		*
NM_004429	Ephrin B1				* \$
Metalloproteinases - related					
NM_016155	MMP-17/ MT4-MMP				*
NM_021801	MMP26	(*)	(*)	(*)	* \$
NM_001110	ADAM10			*	*
NM_001109	ADAM8		* H		(*)
XM_132370#	ADAM1		*		(*)
NM_003254	TIM1	*	*	*	*
Intracellular cytoskeletal					
NM_001665	rho G	(*)		*	*
NM_006113	VAV3			*	*
NM_002086	GRB2		*	*	(*)
NM_001666	C1		* H		
NM_007124	Utrophin				*
Transcription / nuclear					
NM_030756	Tcf4, DNA269446	(*)	*		*
NM_005252	c-Fos		*	*	*
NM_002592	PCNA			*	*
NM_004060	cyclin G		*		
NM_053056	Cyclin D1		*		(*) \$
NM_003401	XRCC4				*
NM_007149	Zinc finger protein				*
Cell surface adhesion / matrix					
XM_053256	MUC1	*	*	*	*
NM_004363	CEA		(*)		*
NM_002483	NCA				*

NM_006350	Follistatin		* H	(*)	* \$
NM_021101	Claudin 1				* \$
NM_012130	Claudin 14				*
NM_003285	tenascin-R	(*)	*		*
NM_001793	CAD3	(*)		*	*
NM_005076	AXO1		* H		
NM_001843	CONT		* H		
NM_000582	Osteopontin	(*)		*	*
NM_006499	Galectin 8	(*)			*
NM_001711	PGS1 (biglycan)	*	* L		
NM_001466	Frizzled 2				* \$
NM_005545	ISLR				* \$
NM_022763	FLJ23399	(*)		*	*
Vascularization					
NM_020404	TEM1		* H		(*)
NM_001147	Tie2 ligand2		*	*	*
NM_003714	STC-2		* H		(*)
NM_005429	VEGFC		*		(*)
NM_000930	tPA			*	*
NM_001955	Endothelin 1		* H		(*)
NM_000361	Thrombomodulin			(*)	*
NM_001993	TF	(*)	*		*
Channel / transmembrane					
NM_005282	GPR4			*	*
NM_006056	GPR66				*
NM_003058	SLC22A2	(*)	(* H)	*	*
NM_002420	MLSN1				*
NM_000702	ATN2, Na/K transport				*

Genes are associated with the disease states B3, dysplasia (D), BE adjacent to carcinoma (BE-CA), or carcinoma (CA) if present in at least 25% of samples tested. (*) indicates gene expression changes associated with 15-25% of samples.

\$ indicates a target of the Wnt signalling pathway.

The gene expression profiles in Barrett's adenocarcinoma share many similarities with colon tumors. For example, epidermal growth factor receptor (EGFR; previously described in carcinoma in BE) (ak-Kasspooles et al., *Internat. J. Cancer* 54:213-219 (1993), along with other growth factor-related or cell-surface proteins such as Cripto CR1, EPHB2, MUC1, NCA/CEACAM6, CEA (Table 3), are often highly expressed in colon cancer (Ciardiello et al., *Proc. Natl. Acad. Sci. USA* 88:7792-7796 (1991); Liu et al., *Cancer* 94:934-939 (2002); Zimmerman et al., *Proc. Natl. Acad. Sci. USA* 84:2960-2964 (1987); Medina et al., *Cancer Res.* 59:1061-1070 (1999); and Ilantzis et al., *Neoplasia* 4:151-163 (2002)). The sodium channel associated with cystic fibrosis, CFTR, was upregulated in adenocarcinoma and can be detected in some cases of high-grade dysplasia (Table 2). This gene is also overexpressed in colon tumors. Furthermore, there is evidence that several genes listed are targets of Wnt signalling pathways (Table 3) (Tetsu and McCormick, *Nature* 398:422-426 (1999); Miwa et al., *Oncol. Res.* 12:469-476 (2000); Marchenko et al., *Biochem. J.* 363:253-262 (2002); Sagara et al., *Biochem. and Biophys. Res. Comm.* 252:117-122 (1998); Lescher et al., *Dev. Dyn.* 213:440-451 (1998); Willert et al., *BMC Dev. Biol.* 2:1-6 (2002); and Tice et al., *J. Biol. Chem.* 277:14329-14335 (2002)), and it is possible that COX-2, which is implicated in colon cancer as well as adenocarcinoma in Barrett's, is a Wnt pathway target (Howe et al., *Cancer Res.* 59:1572-1577 (1999)). An additional synergistic link is suggested by the recent finding that EGFR is activated by prostaglandin E2, a product of COX-2 (Tsuji et al., *Cell* 93:705-716 (1998); Tsuji et al., *Proc. Natl. Acad. Sci. USA* 94:3336-3340 (1997); and Pai et al., *Nature Med.* 8:289-293 (2002)).

More support for Wnt/beta catenin-like induction comes from the strong induction of transcription factor and TCF4 (TCF7L2) in several dysplasia and adenocarcinoma samples (Figure 3A). Knockout studies in mice indicate that TCF4 is necessary for the maintenance of proliferative crypts in the small intestine, and constitutive activity of TCF4 in APC-deficient human epithelial cells may contribute to their malignant transformation (Korinek et al., *Nature Gen.* 19:379-383 (1998)). Given its role in colon carcinogenesis, TCF4 provides another key link between intestinal metaplasia and carcinoma in BE.

Most genes listed represent known genes, but the novel gene FLJ23399 was one of the genes most consistently observed in adenocarcinoma and patient-matched adjacent BE samples (Figure 3B). Expression in BE adjacent to carcinoma suggests the induction may be epigenetic, or possibly reflect small foci of adenocarcinoma that cannot be identified

histologically. Increased expression of this gene was also discovered herein to be associated with colon tumors, and with metastatic prostate tumors (increased expression with metastasis as compared to primary tumors). Its function is unknown, but the presence of 4 type III fibronectin domains in the putative extracellular region suggest a possible role in cell adhesion and/or cell-matrix interactions.

Barrett's Esophagus-to-Adenocarcinoma Disease Progression:

Despite the difficulties associated with sampling and interpretation, the presence and degree of dysplasia is still the most predictive factor for risk of progression to adenocarcinoma (Miros et al., Gut 32:1441-1446 (1991)). Foci of carcinoma typically appear adjacent to dysplasia, and esophageal resections of high-grade dysplasia frequently contain previously unrecognized adenocarcinoma (Falk et al., Gastrointest. Endosc. 49:170-176 (1999); and Cameron and Carpenter, Am. J. Gastroenterol. 92:586-591 (1997)). In this study, by the time dysplasia was apparent, there was evidence of progressive development toward a gene expression profile similar to a differentiated small intestinal enterocyte (along with a small group of genes representative of other intestinal cell types). A possible key contributing factor is the increased expression of TCF4 with advancing disease. Homozygous disruption of TCF4 in mice results in death shortly after birth, and the neonatal epithelium is composed only of non-dividing villus cells (Korinek, V. et al., Nature Gen. 19:379-383 (1998)). This suggests that the genetic program controlled by TCF4 maintains, and possibly establishes, the crypt stem cells of the small intestine. In humans, TCF4 is expressed strongly in the crypts in early fetal development, with increasing expression on the villi up to week 22 as the small intestine develops (Barker et al., Am. J. Pathol. 154:29-35 (1999)). TCF4 is also expressed along the crypt-villus axis of adult small intestine and along the epithelial lining of the crypts of adult colon. The TCF4 profile observed in dysplasia and carcinoma in BE may reflect the inappropriate activation of a developmental pathway with a possible underlying dynamic and differentiating stem cell-like population, or acquisition of some of these characteristics. The delicate cells of the small intestine, with their specialized absorptive and digestive functions and rapid turnover, would seem highly susceptible to damage in the context of the esophagus and gastrointestinal reflux disease.

The developing intestinal phenotype apparent by progression to dysplasia, associated with increased expression of TCF4, suggests some tantalizing links to the development of

carcinoma and the similarities in gene expression between adenocarcinoma of the esophagus and colon. In the context of loss of APC function, association of beta catenin with TCF4 results in constitutive transcription of Tcf target genes, a proposed crucial event in the early transformation of colonic epithelia in colon cancer (Korinek et al., *Science* 275:1784-1787 (1997)). While there is not strong evidence of truncating mutations in APC or oncogenic beta catenin in esophageal adenocarcinoma, there is evidence of hypermethylation of the APC promoter (in 48/52 of adenocarcinoma patients and 17/43 patients with BE metaplasia) (Kawakami et al., *J. Natl. Cancer Inst.* 92:1805-1811 (2000)). APC hypermethylation has also been implicated in progression in colon cancer (Hiltunen et al., *Int. J. Cancer* 70:644-648 (1997)). In this context, it is interesting to note that elevated c-Fos expression was apparent in our study in both dysplasia and carcinoma (Table 3). This could perhaps be related to the presence of bile acids from reflux, overexpression of proglucagon-derived peptide GLP2 (Table 2), or of TNFa (Table 1), all of which have been shown to induce c-Fos expression (Bakin and Curran, *Science* 283:387-390 (1999); Di Toro et al., *Eur. J. Pharm. Sci.* 11:291-298 (2000); and Bjerknes and Cheng, *Proc. Natl. Acad. Sci. USA* 98:12497-12502 (2001)). One proposal for oncogenic transformation by c-Fos is hypermethylation resulting from induction of DNA 5-methylcytosine transferase (Goetze et al., *Atherosclerosis* 159:93-101 (2001)). These factors may contribute to a potential increased availability of beta catenin to combine with TCF4 and activate transcriptional pathways that contribute to carcinogenesis. c-Fos may play an earlier role in intestinal metaplasia as well: studies of intestinal development in mice indicate that GLP2-mediated induction of c-Fos in enteric neurons signals growth of columnar epithelial cell progenitors and stem cells (Di Toro et al., *Eur. J. Pharm. Sci.* 11:291-298 (2000)).

Gene expression profiling of esophageal biopsies has revealed several intriguing associations for the progression of malignancy in the context of Barrett's esophagus. Many of the genes may be involved in potentiating regulatory cycles, and there is potential synergy for the development of adenocarcinoma between exposure to damaging agents (eg. bile), inflammatory response and prostaglandin synthesis, intestinal metaplasia and TCF4 induction, along with induction of growth factors such as EGFR and oncogenes such as c-Fos. Subsets of the genes identified may also eventually serve as markers to identify patients at higher risk for adenocarcinoma. This could permit streamlining of expensive and time-consuming surveillance programs, along with earlier detection and associated improved survival chances for high-risk patients.

Diagnosis of High-grade Esophageal Dysplasia and Prognosis of Esophageal Adenocarcinoma:

5 Several HGD gene markers were discovered as being up-regulated at least 1.5-fold in many high-grade dysplasia samples but are up-regulated in relatively few Barrett's esophagus samples (see Table 4A compared to Table 4B). According to the invention, where at least eight of the twenty-two HGD gene markers are detected to be up-regulated at 1.5-fold in an esophageal tissue sample, cells of the tissue sample are said to exhibit HGD. In addition, the
10 patient from whom the sample was taken may be diagnosed as experiencing high-grade esophageal dysplasia. Further, the prognosis for the patient includes the likely development of adenocarcinoma. Based on the detection of HGD, diagnosis and prognosis, the patient may be treated accordingly and at an earlier stage in the BE-to-cancer progression than would otherwise have occurred prior to disclosure of the instant invention. Alternatively, in a test
15 esophageal tissue sample, where at least one of the at least eight up-regulated HGD marker genes is AGR2 (SEQ ID NO:3), TM7SF1 (SEQ ID NO:13), MAT2B (SEQ ID NO:17), SLNAC1 (SEQ ID NO:23), or TCF4 (SEQ ID NO:43), cells of the tissue sample exhibit HGD and the the patient is said to be diagnosed as experiencing dysplasia, particularly high-grade dysplasia, and is likely to develop adenocarcinoma.

Table 4A High-grade Dysplasia Markers

NCBI #	SEQ ID NO: (na and aa)	Gene name					Sample ID #		
							Z score*		
NM_001955	1 and 2	Endothelin 1	2493	2955	2491	2958	3128	2493	3130
NM_006408	3 and 4	anterior gradient 2 (Xenopus laevis) homolog	2.9		1.9	2.7	2.2		
NM_001109	5 and 6	ADAM8	3.1	2.7	2.6	2.7	3.4	2.	2.9
NM_002773	7 and 8	Prostasin precursor, serine protease	3.6		1.8		2.3		
NM_005076	9 and 10	Axonin-1 precursor	2.5	1.8	2.7		3.1	2.3	
NM_021969	11 and 12	Nuclear hormone receptor	2.		1.6	2.		1.5	
NM_003272	13 and 14	TM7SF1	4.9		2.1	2.8	3.6	2.6	2.7
NM_000108	15 and 16	dihydrolipamide dehydrogenase	1.5	3.6	2.3	1.7	3.	2.2	1.7
NM_013283	17 and 18	methionine adenosyltransferase II, beta	2.1	3.2	1.9	1.7			
NM_003714	19 and 20	stanniocalcin-2	2.5	1.8	2.2	3.	2.7		1.9
NM_001631	21 and 22	Alkaline phosphatase, intestinal precursor	2.3		1.7	1.9	1.6		
NM_004769	23 and 24	Sodium channel receptor SLNAC1	2.3		1.6	2.	2.4	ND	
NM_000717	25 and 26	Carbonic anhydrase iv precursor	2.9	1.8	3.6	3.	2.9	ND	2.5
NM_000928	27 and 28	Phospholipase a2 precursor				1.7	1.8		1.8
NM_005242	29 and 30	Proteinase activated receptor 2 precursor	2.			2.9	2.4	2.4	
NM_004969	31 and 32	Insulin-degrading enzyme		1.6	2.5	4.4		2.7	1.8
NM_005379	33 and 34	Myosin IA (MYO1A)		1.8	2.3	1.5	1.8	1.9	

NM_000775	35 and 36	Cytochrome P450 monooxygenase CYP2J2	CYP2J2		2.4	4.3	2.3		
NM_006214	37 and 38	Phytanoyl-CoA hydroxylase (Refsum disease)	PHYH		2.9	2.4		1.9	
NM_001914	39 and 40	"Cytochrome b5 , 3' end"	CYB5			3.		2.4	
NM_001863	41 and 42	"CoxVlb gene, last exon and flanking sequence"	coxVlb	1.9		2.2	1.9		1.6
NM_030756	43 and 44	TCF4	TCF4	3.6		2.6	3.5	4.1	
		total number		15	10	17	16	12	8

Z score cut-off was 1.5 or above ($p < 0.07$). "na" and "aa" refer to the nucleic acid and amino acid SEQ ID NO, respectively, for the associated markers.

Table 4B Low Prevalence of HGD Markers

NCBI #	SEQ ID NO: (na and aa)	Gene name	Sample ID #																	
			Z score*																	
			B-15	B-17	B-18	B	3091	3131	3132	3142	3143	3088	2296	2554	2555	3134	3135	3140	3181	3141
NM_001955	1 and 2	ET-1																		
NM_006408	3 and 4	AGR2			2.5													1.5		
NM_001109	5 and 6	ADAM8	2.2																	
NM_002773	7 and 8	PRSS8			3.4	1.5														
NM_005076	9 and 10	AXO1																		
NM_021969	11 and 12	NROB2			3.2			2.4	2.4	2.2		1.7		1.7	2.6					
NM_003272	13 and 14	TM7SF1			3.1															
NM_000108	15 and 16	DLDH	2.																	
NM_013283	17 and 18	MAT2B				2.4														
NM_003714	19 and 20	STC-2																		
NM_001631	21 and 22	PPBI						2.												
NM_004769	23 and 24	SLNAC1	2.8																	
NM_000717	25 and 26	CAH4	1.8	1.5							4.2	4.7		2.6	4.3			7.4	1.5	
NM_000928	27 and 28	PA21																		
NM_005242	29 and 30	PAR2																		
NM_004969	31 and 32	IDE				1.5									2.6			2.8	4.9	

[illegible]

Z score cut-off was 1.5 or above ($p < 0.07$). “na” and “aa” refer to the nucleic acid and amino acid SEQ ID NO, respectively, for the associated markers.

In addition to detecting and diagnosing HGD and developing a prognosis of esophageal adenocarcinoma, treatment of cancer, including, but not limited to adenocarcinoma, esophageal adenocarcioma, and colon cancer is also possible by administering to a patient a therapeutically effective amount of an antagonist of one or more of

5 the following adenocarcinoma marker polypeptides: CAD17 (liver-intestine cadherin, NM_004063) (SEQ ID NO:46), CLDN15 (claudin 15, NM_014343) (SEQ ID NO:48), SLNAC1 (sodium channel, NM_004769) (SEQ ID NO:24), CFTR (chloride channel, NM_000492) (SEQ ID NO:50), H2R (histamine H2 receptor, NM_022304) (SEQ ID NO:52), PRSS8 (serine protease, NM_002773) (SEQ ID NO:8), PA21 (phospholipase A2 group IB, 10 NM_000928) (SEQ ID NO:28), AGR2 (anterior gradient 2 homolog, (NM_006408) (SEQ ID NO:4), EGFR (NM_005228) (SEQ ID NO:54), EPHB2 (NM_004442) (SEQ ID NO:56), CRIPTO CR-1 (NM_003212) (SEQ ID NO:58), Eprin B1 (NM_004429) (SEQ ID NO:60), MMP-17/MT4-MMP (NM_016155) (SEQ ID NO:62), MMP26 (NM_021801) (SEQ ID NO:64), ADAM10 (NM_001110) (SEQ ID NO:66), ADAM8 (NM_001109) (SEQ ID NO:6), 15 ADAM1 (XM_132370) (SEQ ID NO:68), TIM1 (NM_003254) (SEQ ID NO:70), MUC1 (XM_053256) (SEQ ID NO:72), CEA (NM_004363) (SEQ ID NO:74), NCA (NM_002483) (SEQ ID NO:76), Follistatin (NM_006350) (SEQ ID NO:78), Claudin 1 (NM_021101) (SEQ ID NO:80), Claudin 14 (NM_012130) (SEQ ID NO:82), tenascin-R (NM_003285) (SEQ ID NO:84), CAD3 (NM_001793) (SEQ ID NO:86), AXO1 (NM_005076) (SEQ ID NO:10), 20 CONT (NM_001843) (SEQ ID NO:88), Osteopontin (NM_000582) (SEQ ID NO:90), Galectin 8 (NM_006499) (SEQ ID NO:92), PGS1 (bihlycan, NM_001711) (SEQ ID NO:94), Frizzled 2 (NM_001466) (SEQ ID NO:96), ISLR (NM_005545) (SEQ ID NO:98), FLJ23399 (NM_022763) (SEQ ID NO:100), TEM1 (NM_020404) (SEQ ID NO:102), Tie2 ligand2 (NM_001147) (SEQ ID NO:104), STC-2 (NM_003714) (SEQ ID NO:20), VEGFC 25 (NM_005429) (SEQ ID NO:106), tPA (NM_000930) (SEQ ID NO:108), Endothelin 1 (NM_001955) (SEQ ID NO:2), Thrombomodulin (NM_000361) (SEQ ID NO:110), TF (NM_001993) (SEQ ID NO:112), GPR4 (NM_005282) (SEQ ID NO:114), GPR66 (NM_006056) (SEQ ID NO:116), SLC22A2 (NM_003058) ((SEQ ID NO:118), MLSN1 (NM_002420) (SEQ ID NO:120), or ATN2 (Na/K transport, NM_000702) (SEQ ID NO:122). 30 The antagonist is a small molecule that binds and inactivates the polypeptide; binds and inactivates a precursor of the polypeptide; prevents translation of the polypeptide; prevents its transcription; or the like. Alternatively, the antagonist is an antibody that specifically binds the polypeptide and inhibits or prevents its activity. Where the antagonist is an antibody, the antibody is optionally a monoclonal antibody, a humanized antibody, or a binding fragment

thereof. The treatment involves contacting a cancer cell with an antagonist of at least one of the polypeptides encoded by the adenocarcinoma marker genes listed above, alternatively with an antagonist of at least three, alternatively with at least five, and alternatively with at least eight of the polypeptides encoded by the adenocarcinoma marker genes listed above.

5

Further, a method of screening for a compound that inhibits cancer cell growth or causes the death of a cancer cell, particularly an adenocarcinoma cell, an esophageal adenocarcinoma cell, or a colon cancer cell, is an aspect of the invention. Accordingly, the screening method involves contacting a cancer cell, such as one expressing at least one, three,
 10 five, eight or more of the adenocarcinoma gene markers selected from the group consisting of CAD17 (liver-intestine cadherin, NM_004063) (SEQ ID NO:45), CLDN15 (claudin 15, NM_014343) (SEQ ID NO:47), SLNAC1 (sodium channel, NM_004769) (SEQ ID NO:23), CFTR (chloride channel, NM_000492) (SEQ ID NO:49), H2R (histamine H2 receptor, NM_022304) (SEQ ID NO:51), PRSS8 (serine protease, NM_002773) (SEQ ID NO:7), PA21
 15 (phospholipase A2 group IB, NM_000928) (SEQ ID NO:27), AGR2 (anterior gradient 2 homolog, (NM_006408) (SEQ ID NO:3), EGFR (NM_005228) (SEQ ID NO:53), EPHB2 (NM_004442) (SEQ ID NO:55), CRIPTO CR-1 (NM_003212) (SEQ ID NO:57), Eprin B1 (NM_004429) (SEQ ID NO:59), MMP-17/MT4-MMP (NM_016155) (SEQ ID NO:61), MMP26 (NM_021801) (SEQ ID NO:63), ADAM10 (NM_001110) (SEQ ID NO:65),
 20 ADAM8 (NM_001109) (SEQ ID NO:5), ADAM1 (XM_132370) (SEQ ID NO:67), TIM1 (NM_003254) (SEQ ID NO:69), MUC1 (XM_053256) (SEQ ID NO:71), CEA (NM_004363) (SEQ ID NO:73), NCA (NM_002483) (SEQ ID NO:75), Follistatin (NM_006350) (SEQ ID NO:77), Claudin 1 (NM_021101) (SEQ ID NO:79), Claudin 14 (NM_012130) (SEQ ID NO:81), tenascin-R (NM_003285) (SEQ ID NO:83), CAD3 (NM_001793) (SEQ ID NO:85),
 25 AXO1 (NM_005076) (SEQ ID NO:9), CONT (NM_001843) (SEQ ID NO:87), Osteopontin (NM_000582) (SEQ ID NO:89), Galectin 8 (NM_006499) (SEQ ID NO:91), PGS1 (bilycan, NM_001711) (SEQ ID NO:93), Frizzled 2 (NM_001466) (SEQ ID NO:95), ISLR (NM_005545) (SEQ ID NO:97), FLJ23399 (NM_022763) (SEQ ID NO:99), TEM1 (NM_020404) (SEQ ID NO:101), Tie2 ligand2 (NM_001147) (SEQ ID NO:103), STC-2
 30 (NM_003714) (SEQ ID NO:19), VEGFC (NM_005429) (SEQ ID NO:105), tPA (NM_000930) (SEQ ID NO:107), Endothelin 1 (NM_001955) (SEQ ID NO:1), Thrombomodulin (NM_000361) (SEQ ID NO:109), TF (NM_001993) (SEQ ID NO:111), GPR4 (NM_005282) (SEQ ID NO:113), GPR66 (NM_006056) (SEQ ID NO:115), SLC22A2 (NM_003058) ((SEQ ID NO:117), MLSN1 (NM_002420) (SEQ ID NO:119), and ATN2

(Na/K transport, NM_000702) (SEQ ID NO:121), followed by determining cancer cell growth inhibition or cancer cell death.

Example 5: Nucleic acid and amino acid sequence identity determinations:

As shown below, Table 5 provides the complete source code for the ALIGN-2 sequence comparison computer program. This source code may be routinely compiled for use on a UNIX operating system to provide the ALIGN-2 sequence comparison computer program.

In addition, disclosed herein are hypothetical exemplifications for using the below described method to determine % amino acid sequence identity and % nucleic acid sequence identity using the ALIGN-2 sequence comparison computer program, wherein "PRO" represents the amino acid sequence of a hypothetical HGD marker polypeptide of interest, "Comparison Protein" represents the amino acid sequence of a polypeptide against which the "PRO" polypeptide of interest is being compared, "PRO-DNA" represents a hypothetical HGD marker polypeptide-encoding nucleic acid sequence of interest, "Comparison DNA" represents the nucleotide sequence of a nucleic acid molecule against which the "PRO-DNA" nucleic acid molecule of interest is being compared, "X", "Y", and "Z" each represent different hypothetical amino acid residues and "N", "L" and "V" each represent different hypothetical nucleotides.

Table 5

```

/*
*
* C-C increased from 12 to 15
* Z is average of EQ
* B is average of ND
* match with stop is _M; stop-stop = 0; J (joker) match = 0
*/
#define      _M      -8      /* value of a match with a stop */

int      _day[26][26] = {
/*      A B C D E F G H I J K L M N O P Q R S T U V W X Y Z */

```

```

/* A */ { 2, 0, -2, 0, 0, -4, 1, -1, -1, 0, -1, -2, -1, 0, _M, 1, 0, -2, 1, 1, 0, 0, -6, 0, -3, 0},
/* B */ { 0, 3, -4, 3, 2, -5, 0, 1, -2, 0, 0, -3, -2, 2, _M, -1, 1, 0, 0, 0, 0, -2, -5, 0, -3, 1},
/* C */ {-2, -4, 15, -5, -5, -4, -3, -3, -2, 0, -5, -6, -5, -4, _M, -3, -5, -4, 0, -2, 0, -2, -8, 0, 0, -5},
/* D */ { 0, 3, -5, 4, 3, -6, 1, 1, -2, 0, 0, -4, -3, 2, _M, -1, 2, -1, 0, 0, 0, -2, -7, 0, -4, 2},
5 /* E */ { 0, 2, -5, 3, 4, -5, 0, 1, -2, 0, 0, -3, -2, 1, _M, -1, 2, -1, 0, 0, 0, -2, -7, 0, -4, 3},
/* F */ {-4, -5, -4, -6, -5, 9, -5, -2, 1, 0, -5, 2, 0, -4, _M, -5, -5, -4, -3, -3, 0, -1, 0, 0, 7, -5},
/* G */ { 1, 0, -3, 1, 0, -5, 5, -2, -3, 0, -2, -4, -3, 0, _M, -1, -1, -3, 1, 0, 0, -1, -7, 0, -5, 0},
/* H */ {-1, 1, -3, 1, 1, -2, -2, 6, -2, 0, 0, -2, -2, 2, _M, 0, 3, 2, -1, -1, 0, -2, -3, 0, 0, 2},
/* I */ {-1, -2, -2, -2, -2, 1, -3, -2, 5, 0, -2, 2, 2, -2, _M, -2, -2, -2, -1, 0, 0, 4, -5, 0, -1, -2},
10 /* J */ { 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, _M, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0},
/* K */ {-1, 0, -5, 0, 0, -5, -2, 0, -2, 0, 5, -3, 0, 1, _M, -1, 1, 3, 0, 0, 0, -2, -3, 0, -4, 0},
/* L */ {-2, -3, -6, -4, -3, 2, -4, -2, 2, 0, -3, 6, 4, -3, _M, -3, -2, -3, -3, -1, 0, 2, -2, 0, -1, -2},
/* M */ {-1, -2, -5, -3, -2, 0, -3, -2, 2, 0, 0, 4, 6, -2, _M, -2, -1, 0, -2, -1, 0, 2, -4, 0, -2, -1},
/* N */ { 0, 2, -4, 2, 1, -4, 0, 2, -2, 0, 1, -3, -2, 2, _M, -1, 1, 0, 1, 0, 0, -2, -4, 0, -2, 1},
15 /* O */ { _M, _M, _M, _M, _M, _M, _M, _M, _M, _M, _M, _M, _M,
0, _M, _M, _M, _M, _M, _M, _M, _M, _M, _M,
/* P */ { 1, -1, -3, -1, -1, -5, -1, 0, -2, 0, -1, -3, -2, -1, _M, 6, 0, 0, 1, 0, 0, -1, -6, 0, -5, 0},
/* Q */ { 0, 1, -5, 2, 2, -5, -1, 3, -2, 0, 1, -2, -1, 1, _M, 0, 4, 1, -1, -1, 0, -2, -5, 0, -4, 3},
/* R */ {-2, 0, -4, -1, -1, -4, -3, 2, -2, 0, 3, -3, 0, 0, _M, 0, 1, 6, 0, -1, 0, -2, 2, 0, -4, 0},
20 /* S */ { 1, 0, 0, 0, 0, -3, 1, -1, -1, 0, 0, -3, -2, 1, _M, 1, -1, 0, 2, 1, 0, -1, -2, 0, -3, 0},
/* T */ { 1, 0, -2, 0, 0, -3, 0, -1, 0, 0, 0, -1, -1, 0, _M, 0, -1, -1, 1, 3, 0, 0, -5, 0, -3, 0},
/* U */ { 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, _M, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0},
/* V */ { 0, -2, -2, -2, -2, -1, -1, -2, 4, 0, -2, 2, 2, -2, _M, -1, -2, -2, -1, 0, 0, 4, -6, 0, -2, -2},
/* W */ {-6, -5, -8, -7, -7, 0, -7, -3, -5, 0, -3, -2, -4, -4, _M, -6, -5, 2, -2, -5, 0, -6, 17, 0, 0, -6},
25 /* X */ { 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, _M, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0},
/* Y */ {-3, -3, 0, -4, -4, 7, -5, 0, -1, 0, -4, -1, -2, -2, _M, -5, -4, -4, -3, -3, 0, -2, 0, 0, 10, -4},
/* Z */ { 0, 1, -5, 2, 3, -5, 0, 2, -2, 0, 0, -2, -1, 1, _M, 0, 3, 0, 0, 0, 0, -2, -6, 0, -4, 4}
};

```


5

10

Page 1 of day.h

/*

*/

#include <stdio.h>

#include <ctype.h>

15

#define MAXJMP 16 /* max jumps in a diag */

#define MAXGAP 24 /* don't continue to penalize gaps larger than this */

#define JMPS 1024 /* max jmps in an path */

#define MX 4 /* save if there's at least MX-1 bases since last jmp */

20

#define DMAT 3 /* value of matching bases */

#define DMIS 0 /* penalty for mismatched bases */

#define DINS0 8 /* penalty for a gap */

#define DINS1 1 /* penalty per base */

25

#define PINS0 8 /* penalty for a gap */

#define PINS1 4 /* penalty per residue */

struct jmp {

short n[MAXJMP]; /* size of jmp (neg for dely) */

30

unsigned short x[MAXJMP]; /* base no. of jmp in seq x */

}; /* limits seq to 2¹⁶ -1 */

struct diag {

int score; /* score at last jmp */

```

        long      offset;      /* offset of prev block */
        short     ijmp;        /* current jmp index */
        struct jmp jp;         /* list of jmps */
};

5
struct path {
    int    spc;                /* number of leading spaces */
    short  n[JMPs];           /* size of jmp (gap) */
    int    x[JMPs];           /* loc of jmp (last elem before gap) */
10 };

    char      *ofile;          /* output file name */
    char      *namex[2];        /* seq names: getseqs() */
    char      *prog;            /* prog name for err msgs */
15  char      *seqx[2];         /* seqs: getseqs() */
    int       dmax;             /* best diag: nw() */
    int       dmax0;            /* final diag */
    int       dna;              /* set if dna: main() */
    int       endgaps;           /* set if penalizing end gaps */
20  int       gapx, gapy;        /* total gaps in seqs */
    int       len0, len1;        /* seq lens */
    int       ngapx, ngapy;      /* total size of gaps */
    int       smax;             /* max score: nw() */
    int       *xbm;             /* bitmap for matching */
25  long      offset;           /* current offset in jmp file */
    struct diag *dx;            /* holds diagonals */
    struct path pp[2];          /* holds path for seqs */

    char      *calloc(), *malloc(), *index(), *strcpy();
30  char      *getseq(), *g_calloc();

```

```

/* Needleman-Wunsch alignment program
*
* usage: progs file1 file2
5  * where file1 and file2 are two dna or two protein sequences.
* The sequences can be in upper- or lower-case and may contain ambiguity
* Any lines beginning with ';', '>' or '<' are ignored
* Max file length is 65535 (limited by unsigned short x in the jmp struct)
* A sequence with 1/3 or more of its elements ACGTU is assumed to be DNA
10 * Output is in the file "align.out"
*
* The program may create a tmp file in /tmp to hold info about traceback.
* Original version developed under BSD 4.3 on a vax 8650
*/
15 #include "nw.h"
#include "day.h"

static _dbval[26] = {
    1,14,2,13,0,0,4,11,0,0,12,0,3,15,0,0,0,5,6,8,8,7,9,0,10,0
20 };

static _pbval[26] = {
    1, 2|(1<<('D'-'A'))|(1<<('N'-'A')), 4, 8, 16, 32, 64,
    128, 256, 0xFFFFFFFF, 1<<10, 1<<11, 1<<12, 1<<13, 1<<14,
25 1<<15, 1<<16, 1<<17, 1<<18, 1<<19, 1<<20, 1<<21, 1<<22,
    1<<23, 1<<24, 1<<25|(1<<('E'-'A'))|(1<<('Q'-'A'))
};

main(ac, av)
30     int    ac;
    char    *av[];
{
    prog = av[0];
    if (ac != 3) {

```

main

```
    fprintf(stderr,"usage: %s file1 file2\n", prog);
    fprintf(stderr,"where file1 and file2 are two dna or two protein sequences.\n");
    fprintf(stderr,"The sequences can be in upper- or lower-case\n");
    fprintf(stderr,"Any lines beginning with ';' or '<' are ignored\n");
5    fprintf(stderr,"Output is in the file \"align.out\"\n");
    exit(1);
}
    namex[0] = av[1];
    namex[1] = av[2];
10    seqx[0] = getseq(namex[0], &len0);
    seqx[1] = getseq(namex[1], &len1);
    xbm = (dna)? _dbval : _pbval;

    endgaps = 0;                /* 1 to penalize endgaps */
15    ofile = "align.out";       /* output file */

    nw();                       /* fill in the matrix, get the possible jumps */
    readjumps();                /* get the actual jumps */
    print();                     /* print stats, alignment */
20
    cleanup(0);                 /* unlink any tmp files */
}
```

Page 1 of nw.c

```

/* do the alignment, return best score: main()
* dna: values in Fitch and Smith, PNAS, 80, 1382-1386, 1983
* pro: PAM 250 values
5  * When scores are equal, we prefer mismatches to any gap, prefer
* a new gap to extending an ongoing gap, and prefer a gap in seqx
* to a gap in seq y.
*/
nw()
10 {
    char      *px, *py;          /* seqs and ptrs */
    int        *ndely, *dely; /* keep track of dely */
    int        ndelx, delx;    /* keep track of delx */
    int        *tmp;          /* for swapping row0, row1 */
    15 int        mis;          /* score for each type */
    int        ins0, ins1;    /* insertion penalties */
    register    id;           /* diagonal index */
    register    ij;           /* jmp index */
    register    *col0, *col1; /* score for curr, last row */
    20 register    xx, yy;      /* index into seqs */

    dx = (struct diag *)g_calloc("to get diags", len0+len1+1, sizeof(struct diag));

    ndely = (int *)g_calloc("to get ndely", len1+1, sizeof(int));
    25 dely = (int *)g_calloc("to get dely", len1+1, sizeof(int));
    col0 = (int *)g_calloc("to get col0", len1+1, sizeof(int));
    col1 = (int *)g_calloc("to get col1", len1+1, sizeof(int));
    ins0 = (dna)? DINS0 : PINS0;
    ins1 = (dna)? DINS1 : PINS1;

    30 smax = -10000;
    if (endgaps) {
        for (col0[0] = dely[0] = -ins0, yy = 1; yy <= len1; yy++) {
            col0[yy] = dely[yy] = col0[yy-1] - ins1;
        }
    }
}

```

```
        ndely[yy] = yy;
    }
    col0[0] = 0;    /* Waterman Bull Math Biol 84 */
}
5   else
        for (yy = 1; yy <= len1; yy++)
            dely[yy] = -ins0;

/* fill in match matrix
10  */
    for (px = seqx[0], xx = 1; xx <= len0; px++, xx++) {
        /* initialize first entry in col
        */
        if (endgaps) {
15             if (xx == 1)
                    col1[0] = delx = -(ins0+ins1);
                else
                    col1[0] = delx = col0[0] - ins1;
                ndelx = xx;
20         }
        else {
            col1[0] = 0;
            delx = -ins0;
            ndelx = 0;
25         }
    }
```

...nw

```

for (py = seqx[1], yy = 1; yy <= len1; py++, yy++) {
    mis = col0[yy-1];
5    if (dna)
        mis += (xbm[*px-'A']&xbm[*py-'A'])? DMAT : DMIS;
    else
        mis += _day[*px-'A'][*py-'A'];

10    /* update penalty for del in x seq;
        * favor new del over ongong del
        * ignore MAXGAP if weighting endgaps
        */
    if (endgaps || ndely[yy] < MAXGAP) {
15        if (col0[yy] - ins0 >= dely[yy]) {
            dely[yy] = col0[yy] - (ins0+ins1);
            ndely[yy] = 1;
        } else {
            dely[yy] -= ins1;
20            ndely[yy]++;
        }
    } else {
        if (col0[yy] - (ins0+ins1) >= dely[yy]) {
            dely[yy] = col0[yy] - (ins0+ins1);
25            ndely[yy] = 1;
        } else
            ndely[yy]++;
    }

30    /* update penalty for del in y seq;
        * favor new del over ongong del
        */
    if (endgaps || ndelx < MAXGAP) {
        if (col1[yy-1] - ins0 >= delx) {

```

```

        delx = col1[yy-1] - (ins0+ins1);
        ndelx = 1;
    } else {
        delx -= ins1;
5       ndelx++;
    }
} else {
    if (col1[yy-1] - (ins0+ins1) >= delx) {
        delx = col1[yy-1] - (ins0+ins1);
10       ndelx = 1;
    } else
        ndelx++;
}

15  /* pick the maximum score; we're favoring
    * mis over any del and delx over dely
    */

```

20

25

...nw

```

id = xx - yy + len1 - 1;
if (mis >= delx && mis >= dely[yy])
    col1[yy] = mis;
5  else if (delx >= dely[yy]) {
    col1[yy] = delx;
    ij = dx[id].ijmp;
    if (dx[id].jp.n[0] && (!dna || (ndelx >= MAXJMP
10  && xx > dx[id].jp.x[ij]+MX) || mis > dx[id].score+DINS0)) {
        dx[id].ijmp++;
        if (++ij >= MAXJMP) {
            writejumps(id);
            ij = dx[id].ijmp = 0;
            dx[id].offset = offset;
15  offset += sizeof(struct jmp) + sizeof(offset);
        }
    }
    dx[id].jp.n[ij] = ndelx;
    dx[id].jp.x[ij] = xx;
20  dx[id].score = delx;
}
else {
    col1[yy] = dely[yy];
    ij = dx[id].ijmp;
25
    if (dx[id].jp.n[0] && (!dna || (ndely[yy] >= MAXJMP
        && xx > dx[id].jp.x[ij]+MX) || mis > dx[id].score+DINS0)) {
        dx[id].ijmp++;
        if (++ij >= MAXJMP) {
30  writejumps(id);
            ij = dx[id].ijmp = 0;
            dx[id].offset = offset;
            offset += sizeof(struct jmp) + sizeof(offset);
        }
    }
}

```

```

        }
        dx[id].jp.n[ij] = -ndely[yy];
        dx[id].jp.x[ij] = xx;
        dx[id].score = dely[yy];
5      }
      if (xx == len0 && yy < len1) {
        /* last col
        */
        if (endgaps)
10          col1[yy] -= ins0+ins1*(len1-yy);
        if (col1[yy] > smax) {
          smax = col1[yy];
          dmax = id;
        }
15      }
    }
    if (endgaps && xx < len0)
        col1[yy-1] -= ins0+ins1*(len0-xx);
    if (col1[yy-1] > smax) {
20        smax = col1[yy-1];
        dmax = id;
    }
    tmp = col0; col0 = col1; col1 = tmp;
}
25 (void) free((char *)ndely);
(void) free((char *)dely);
(void) free((char *)col0);
(void) free((char *)col1);
}

```

30

Page 4 of nw.c

```

/*
*
* print() -- only routine visible outside this module
5  *
* static:
* getmat() -- trace back best path, count matches: print()
* pr_align() -- print alignment of described in array p[]: print()
* dumpblock() -- dump a block of lines with numbers, stars: pr_align()
10 * nums() -- put out a number line: dumpblock()
* putline() -- put out a line (name, [num], seq, [num]): dumpblock()
* stars() - -put a line of stars: dumpblock()
* stripname() -- strip any path and prefix from a seqname
*/
15
#include "nw.h"

#define SPC 3
#define P_LINE 256 /* maximum output line */
20 #define P_SPC 3 /* space between name or num and seq */

extern _day[26][26];
int olen; /* set output line length */
FILE *fx; /* output file */
25

print() print
{
    int lx, ly, firstgap, lastgap; /* overlap */

30    if ((fx = fopen(ofile, "w")) == 0) {
        fprintf(stderr, "%s: can't write %s\n", prog, ofile);
        cleanup(1);
    }
    fprintf(fx, "<first sequence: %s (length = %d)\n", namex[0], len0);

```

```
fprintf(fx, "<second sequence: %s (length = %d)\n", namex[1], len1);
olen = 60;
lx = len0;
ly = len1;
5 firstgap = lastgap = 0;
  if (dmax < len1 - 1) { /* leading gap in x */
    pp[0].spc = firstgap = len1 - dmax - 1;
    ly -= pp[0].spc;
  }
10 else if (dmax > len1 - 1) { /* leading gap in y */
    pp[1].spc = firstgap = dmax - (len1 - 1);
    lx -= pp[1].spc;
  }
  if (dmax0 < len0 - 1) { /* trailing gap in x */
15 lastgap = len0 - dmax0 - 1;
    lx -= lastgap;
  }
  else if (dmax0 > len0 - 1) { /* trailing gap in y */
    lastgap = dmax0 - (len0 - 1);
20 ly -= lastgap;
  }
  getmat(lx, ly, firstgap, lastgap);
  pr_align();
}
25
```

```

/*
 * trace back the best path, count matches
 */
5  static
    getmat(lx, ly, firstgap, lastgap)                                getmat
        int    lx, ly;                /* "core" (minus endgaps) */
        int    firstgap, lastgap;      /* leading trailing overlap */
    {
10         int      nm, i0, i1, siz0, siz1;
        char      outx[32];
        double     pct;
        register   n0, n1;
        register char *p0, *p1;

15         /* get total matches, score
           */
        i0 = i1 = siz0 = siz1 = 0;
        p0 = seqx[0] + pp[1].spc;
20         p1 = seqx[1] + pp[0].spc;
        n0 = pp[1].spc + 1;
        n1 = pp[0].spc + 1;

        nm = 0;
25         while ( *p0 && *p1 ) {
            if (siz0) {
                p1++;
                n1++;
                siz0--;
30             }
            else if (siz1) {
                p0++;
                n0++;
                siz1--;

```

```

        }
        else {
            if (xbm[*p0-'A']&xbm[*p1-'A'])
                nm++;
5         if (n0++ == pp[0].x[i0])
                siz0 = pp[0].n[i0++];
            if (n1++ == pp[1].x[i1])
                siz1 = pp[1].n[i1++];
            p0++;
10         p1++;
        }
    }

    /* pct homology:
15    * if penalizing endgaps, base is the shorter seq
    * else, knock off overhangs and take shorter core
    */
    if (endgaps)
        lx = (len0 < len1)? len0 : len1;
20    else
        lx = (lx < ly)? lx : ly;
    pct = 100.*(double)nm/(double)lx;
    fprintf(fx, "\n");
    fprintf(fx, "<%d match%s in an overlap of %d: %.2f percent similarity\n",
25        nm, (nm == 1)? "" : "es", lx, pct);

```

```

    fprintf(fx, "<gaps in first sequence: %d", gapx);
    if (gapx) {
        (void) sprintf(outx, " (%d %s%s)",
5          ngapx, (dna)? "base":"residue", (ngapx == 1)? "" : "s");
        fprintf(fx, "%s", outx);

    fprintf(fx, ", gaps in second sequence: %d", gapy);
    if (gapy) {
10      (void) sprintf(outx, " (%d %s%s)",
          ngapy, (dna)? "base":"residue", (ngapy == 1)? "" : "s");
        fprintf(fx, "%s", outx);
    }
    if (dna)
15      fprintf(fx,
        "\n<score: %d (match = %d, mismatch = %d, gap penalty = %d + %d per
base)\n",
        smax, DMAT, DMIS, DINS0, DINS1);
    else
20      fprintf(fx,
        "\n<score: %d (Dayhoff PAM 250 matrix, gap penalty = %d + %d per
residue)\n",
        smax, PINS0, PINS1);
    if (endgaps)
25      fprintf(fx,
        "<endgaps penalized. left endgap: %d %s%s, right endgap: %d %s%s\n",
        firstgap, (dna)? "base" : "residue", (firstgap == 1)? "" : "s",
        lastgap, (dna)? "base" : "residue", (lastgap == 1)? "" : "s");
    else
30      fprintf(fx, "<endgaps not penalized\n");
}

static      nm;          /* matches in core -- for checking */
static      lmax;        /* lengths of stripped file names */

```

...getmat

```

static      ij[2];          /* jmp index for a path */
static      nc[2];          /* number at start of current line */
static      ni[2];          /* current elem number -- for gapping */
static      siz[2];
5  static char *ps[2];       /* ptr to current element */
static char *po[2];         /* ptr to next output char slot */
static char out[2][P_LINE]; /* output line */
static char star[P_LINE]; /* set by stars() */

10 /*
   * print alignment of described in struct path pp[]
   */
static
pr_align()                                pr_align
15 {
    int      nn;      /* char count */
    int      more;
    register i;

20     for (i = 0, lmax = 0; i < 2; i++) {
        nn = stripname(namex[i]);
        if (nn > lmax)
            lmax = nn;

25         nc[i] = 1;
        ni[i] = 1;
        siz[i] = ij[i] = 0;
        ps[i] = seqx[i];
        po[i] = out[i];

30     }

```


...pr_align

```

for (nn = nm = 0, more = 1; more; ) {
    for (i = more = 0; i < 2; i++) {
        /*
5         * do we have more of this sequence?
        */
        if (!*ps[i])
            continue;

10        more++;

        if (pp[i].spc) { /* leading space */
            *po[i]++ = ' ';
            pp[i].spc--;
15        }
        else if (siz[i]) { /* in a gap */
            *po[i]++ = '-';
            siz[i]--;
        }
20        else { /* we're putting a seq element
            */
            *po[i] = *ps[i];
            if (islower(*ps[i]))
                *ps[i] = toupper(*ps[i]);
25            po[i]++;
            ps[i]++;

            /*
            * are we at next gap for this seq?
            */
30            if (ni[i] == pp[i].x[ij[i]]) {
                /*
                * we need to merge all gaps
                * at this location

```

```

        */
        siz[i] = pp[i].n[ij[i]++];
        while (ni[i] == pp[i].x[ij[i]])
            siz[i] += pp[i].n[ij[i]++];
5          }
          ni[i]++;
        }
    }
    if (++nn == olen || !more && nn) {
10        dumpblock();
        for (i = 0; i < 2; i++)
            po[i] = out[i];
        nn = 0;
    }
15 }
}

/*
 * dump a block of lines, including numbers, stars: pr_align()
20 */
static
dumpblock()                                dumpblock
{
    register    i;
25
    for (i = 0; i < 2; i++)
        *po[i]-- = '\0';

```

...dumpblock

```

    (void) putc('\n', fx);
5    for (i = 0; i < 2; i++) {
        if (*out[i] && (*out[i] != ' ' || *(po[i]) != ' ')) {
            if (i == 0)
                nums(i);
            if (i == 0 && *out[1])
10                stars();
            putline(i);
            if (i == 0 && *out[1])
                fprintf(fx, star);
            if (i == 1)
15                nums(i);
        }
    }
}

```

```

20  /*
    * put out a number line: dumpblock()
    */
    static
    nums(ix)
25    int    ix;    /* index in out[] holding seq line */
    {
        char    nline[P_LINE];
        register    i, j;
        register char    *pn, *px, *py;
30
        for (pn = nline, i = 0; i < lmax+P_SPC; i++, pn++)
            *pn = ' ';
        for (i = nc[ix], py = out[ix]; *py; py++, pn++) {
            if (*py == ' ' || *py == '-')

```

nums

```

        *pn = ' ';
    else {
        if (i%10 == 0 || (i == 1 && nc[ix] != 1)) {
            j = (i < 0)? -i : i;
            for (px = pn; j /= 10, px--)
                *px = j%10 + '0';
            if (i < 0)
                *px = '-';
        }
        else
            *pn = ' ';
        i++;
    }
}

*pn = '\0';
nc[ix] = i;
for (pn = nline; *pn; pn++)
    (void) putc(*pn, fx);
(void) putc('\n', fx);
}

/*
 * put out a line (name, [num], seq, [num]): dumpblock()
 */
static
putline(ix)                                putline
    int    ix;
{

```

...putline

```

    int            i;
5    register char *px;

    for (px = namex[ix], i = 0; *px && *px != ':'; px++, i++)
        (void) putc(*px, fx);
    for (; i < lmax+P_SPC; i++)
10        (void) putc(' ', fx);

    /* these count from 1:
       * ni[] is current element (from 1)
       * nc[] is number at start of current line
15    */
    for (px = out[ix]; *px; px++)
        (void) putc(*px&0x7F, fx);
    (void) putc('\n', fx);
}
20

/*
 * put a line of stars (seqs always in out[0], out[1]): dumpblock()
 */
25 static
stars()
{
    int            i;
    register char *p0, *p1, cx, *px;
30

    if (!*out[0] || (*out[0] == ' ' && *(po[0]) == ' ') ||
        !*out[1] || (*out[1] == ' ' && *(po[1]) == ' '))
        return;
    px = star;

```

stars

```

    for (i = lmax+P_SPC; i; i--)
        *px++ = ' ';

    for (p0 = out[0], p1 = out[1]; *p0 && *p1; p0++, p1++) {
5        if (isalpha(*p0) && isalpha(*p1)) {

            if (xbm[*p0-'A']&xbm[*p1-'A']) {
                cx = '*';
                nm++;
10            }
            else if (!dna && _day[*p0-'A'][*p1-'A'] > 0)
                cx = '!';
            else
                cx = ' ';
15        }
        else
            cx = ' ';
        *px++ = cx;
    }
20    *px++ = '\n';
    *px = '\0';
}

25
```

```
/*
 * strip path or prefix from pn, return len: pr_align()
 */
5  static
stripname(pn)                                stripname
    char *pn; /* file name (may be path) */
{
    register char *px, *py;
10
    py = 0;
    for (px = pn; *px; px++)
        if (*px == '/')
            py = px + 1;
15    if (py)
        (void) strcpy(pn, py);
    return(strlen(pn));
}
20

25

30
```

```

/*
 * cleanup() -- cleanup any tmp file
 * getseq() -- read in seq, set dna, len, maxlen
5  * g_calloc() -- calloc() with error checkin
 * readjumps() -- get the good jumps, from tmp file if necessary
 * writejumps() -- write a filled array of jumps to a tmp file: nw()
 */
#include "nw.h"
10 #include <sys/file.h>

char *jname = "/tmp/homgXXXXXX";          /* tmp file for jumps */
FILE *fj;

15 int cleanup();                          /* cleanup tmp file */
long lseek();

/*
 * remove any tmp file if we blow
20 */
cleanup(i)                                cleanup
    int i;
{
    if (fj)
25     (void) unlink(jname);
    exit(i);
}

/*
30 * read, return ptr to seq, set dna, len, maxlen
 * skip lines starting with ';', '<', or '>'
 * seq in upper or lower case
 */
char *

```



```

getseq(file, len)
char    *file; /* file name */
int     *len;  /* seq len */
{
5   char    line[1024], *pseq;
   register char *px, *py;
   int      natgc, tlen;
   FILE     *fp;

10  if ((fp = fopen(file,"r")) == 0) {
        fprintf(stderr,"%s: can't read %s\n", prog, file);
        exit(1);
    }
    tlen = natgc = 0;
15  while (fgets(line, 1024, fp)) {
        if (*line == ';' || *line == '<' || *line == '>')
            continue;
        for (px = line; *px != '\n'; px++)
            if (isupper(*px) || islower(*px))
20                tlen++;
    }
    if ((pseq = malloc((unsigned)(tlen+6))) == 0) {
        fprintf(stderr,"%s: malloc() failed to get %d bytes for %s\n", prog, tlen+6, file);
        exit(1);
25    }
    pseq[0] = pseq[1] = pseq[2] = pseq[3] = '\0';

```

...getseq

```

py = pseq + 4;
*len = tlen;
5  rewind(fp);

while (fgets(line, 1024, fp)) {
    if (*line == ';' || *line == '<' || *line == '>')
        continue;
10    for (px = line; *px != '\n'; px++) {
        if (isupper(*px))
            *py++ = *px;
        else if (islower(*px))
            *py++ = toupper(*px);
15    if (index("ATGCU",*(py-1)))
        natgc++;
    }
}
*py++ = '\0';
20 *py = '\0';
(void) fclose(fp);
dna = natgc > (tlen/3);
return(pseq+4);
}
25
char *
g_calloc(msg, nx, sz)
    char *msg;        /* program, calling routine */
    int nx, sz;        /* number and size of elements */
30 {
    char *px, *calloc();

    if ((px = calloc((unsigned)nx, (unsigned)sz)) == 0) {
        if (*msg) {

```

g_calloc

```

        fprintf(stderr, "%s: g_calloc() failed %s (n=%d, sz=%d)\n", prog, msg,
nx, sz);

        exit(1);
    }
5      }
    return(px);
}

/*
10  * get final jmps from dx[] or tmp file, set pp[], reset dmax: main()
  */
readjumps()                                readjumps
{
    int      fd = -1;
15    int      siz, i0, i1;
    register  i, j, xx;

    if (fj) {
        (void) fclose(fj);
20    if ((fd = open(jname, O_RDONLY, 0)) < 0) {
        fprintf(stderr, "%s: can't open() %s\n", prog, jname);
        cleanup(1);
    }
}
25  for (i = i0 = i1 = 0, dmax0 = dmax, xx = len0; ; i++) {
    while (1) {
        for (j = dx[dmax].ijmp; j >= 0 && dx[dmax].jp.x[j] >= xx; j--)
            ;

```

Page 2 of nwsubr.c

30

...readjumps

```

    if (j < 0 && dx[dmax].offset && fj) {
        (void) lseek(fd, dx[dmax].offset, 0);
        (void) read(fd, (char *)&dx[dmax].jp, sizeof(struct jmp));
5        (void) read(fd, (char *)&dx[dmax].offset,
        sizeof(dx[dmax].offset));
        dx[dmax].ijmp = MAXJMP-1;
    }
    else
10        break;
}
if (i >= JMPS) {
    fprintf(stderr, "%s: too many gaps in alignment\n", prog);
    cleanup(1);
15 }
if (j >= 0) {
    siz = dx[dmax].jp.n[j];
    xx = dx[dmax].jp.x[j];
    dmax += siz;
20    if (siz < 0) {          /* gap in second seq */
        pp[1].n[i1] = -siz;
        xx += siz;

        /* id = xx - yy + len1 - 1
25        */
        pp[1].x[i1] = xx - dmax + len1 - 1;
        gapy++;
        ngapy -= siz;
    /* ignore MAXGAP when doing endgaps */
30    siz = (-siz < MAXGAP || endgaps)? -siz : MAXGAP;
    i1++;
}
else if (siz > 0) {        /* gap in first seq */
    pp[0].n[i0] = siz;

```

```

        pp[0].x[i0] = xx;
        gapx++;
        ngapx += siz;
/* ignore MAXGAP when doing endgaps */
5         siz = (siz < MAXGAP || endgaps)? siz : MAXGAP;
        i0++;
    }
}
else
10         break;
}

/* reverse the order of jmps
*/
15 for (j = 0, i0--; j < i0; j++, i0--) {
    i = pp[0].n[j]; pp[0].n[j] = pp[0].n[i0]; pp[0].n[i0] = i;
    i = pp[0].x[j]; pp[0].x[j] = pp[0].x[i0]; pp[0].x[i0] = i;
}
for (j = 0, i1--; j < i1; j++, i1--) {
20     i = pp[1].n[j]; pp[1].n[j] = pp[1].n[i1]; pp[1].n[i1] = i;
    i = pp[1].x[j]; pp[1].x[j] = pp[1].x[i1]; pp[1].x[i1] = i;
}
if (fd >= 0)
    (void) close(fd);
25 if (fj) {
    (void) unlink(jname);
    fj = 0;
    offset = 0;
}
30 }

```

Page 3 of nwsubr.c

```
/*
 * write a filled jmp struct offset of the prev one (if any): nw()
5  */
writejumps(ix)                                writejumps
    int    ix;
{
    char    *mktemp();
10
    if (!fj) {
        if (mktemp(jname) < 0) {
            fprintf(stderr, "%s: can't mktemp() %s\n", prog, jname);
            cleanup(1);
15
        }
        if ((fj = fopen(jname, "w")) == 0) {
            fprintf(stderr, "%s: can't write %s\n", prog, jname);
            exit(1);
        }
20
    }
    (void) fwrite((char *)&dx[ix].jp, sizeof(struct jmp), 1, fj);
    (void) fwrite((char *)&dx[ix].offset, sizeof(dx[ix].offset), 1, fj);
}

25

30
```

Example calculations for determining % amino acid sequence identity and nucleic acid sequence identity:

1.

PRO XXXXXXXXXXXXXXXXXXXX (Length = 15 amino acids)
 5 Comparison Protein XXXXXYYYYYYYYY (Length = 12 amino acids)

% amino acid sequence identity =

(the number of identically matching amino acid residues between the two polypeptide
 10 sequences as determined by ALIGN-2) divided by (the total number of amino acid residues of the PRO polypeptide) =

5 divided by 15 = 33.3%

15 2.

PRO XXXXXXXXXXXX (Length = 10 amino acids)
 Comparison Protein XXXXXYYYYYYYZZYZ (Length = 15 amino acids)

% amino acid sequence identity =

20

(the number of identically matching amino acid residues between the two polypeptide
 sequences as determined by ALIGN-2) divided by (the total number of amino acid residues of the PRO polypeptide) =

25 5 divided by 10 = 50%

3.

PRO-DNA NNNNNNNNNNNNNNNN (Length = 14 nucleotides)
 Comparison DNA NNNNNNLLLLLLLLLL (Length = 16 nucleotides)

30

% nucleic acid sequence identity =

(the number of identically matching nucleotides between the two nucleic acid sequences as determined by ALIGN-2) divided by (the total number of nucleotides of the PRO-DNA nucleic acid sequence) =

5 6 divided by 14 = 42.9%

4.

PRO-DNA	NNNNNNNNNNNNNN	(Length = 12 nucleotides)
Comparison DNA	NNNNLLL	(Length = 9 nucleotides)

10

% nucleic acid sequence identity =

(the number of identically matching nucleotides between the two nucleic acid sequences as determined by ALIGN-2) divided by (the total number of nucleotides of the PRO-DNA nucleic acid sequence) =

15

4 divided by 12 = 33.3%

20

Although the foregoing refers to particular embodiments, it will be understood that the present invention is not so limited. It will occur to those of ordinary skill in the art that various modifications may be made to the disclosed embodiments without diverting from the overall concept of the invention. All such modifications are intended to be within the scope of the present invention.

25

What is claimed is:

CLAIMS

1. A method of detecting of high-grade dysplasia (HGD) in cells of a tissue sample, the method comprising:

(a) obtaining a test tissue sample suspected of comprising cells exhibiting HGD;

(b) establishing the level of expression in the test tissue sample of at least eight genes selected from the group consisting of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3); ADAM8 (NM_001109) (SEQ ID NO:5); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11); TM7SF1 (NM_003272) (SEQ ID NO:13); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41); and TCF4 (NM_030756) (SEQ ID NO:43), or variants thereof having at least 80% nucleic acid sequence identity, wherein the tissue is from esophagus or colon; and

(c) comparing expression of the at least eight genes to a baseline expression of the genes in normal tissue controls of the same tissue type, wherein an increase of at least 1.5-fold in expression of the genes relative to the baseline expression indicates that cells of the test sample exhibit HGD.

2. The method of claim 1, wherein the tissue is human tissue.

3. A method of identifying a esophageal tissue susceptible to esophageal adenocarcoma, comprising detecting esophageal HGD in a test tissue sample according to claim 1.

4. A method according to claim 1, wherein an increase of at least 2-fold in expression of genes relative to the baseline is observed.

5. A method according to claim 1, wherein at least one of the at least eight genes is selected from the group consisting of AGR2 (SEQ ID NO:3), TM7SF1 (SEQ ID NO:13), MAT2B (SEQ ID NO:17), SLNAC1 (SEQ ID NO:23), and TCF4 (SEQ ID NO:43), or variants thereof having at least 80% nucleic acid sequence identity.

6. A method for determining predisposition of a mammalian tissue to a neo-plastic transformation by detecting HGD in cells of the tissue, the method comprising determining in a cell from the tissue expression of a nucleic acid sequence of at least eight genes selected from the group consisting of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3); ADAM8 (NM_001109) (SEQ ID NO:5); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11); TM7SF1 (NM_003272) (SEQ ID NO:13); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41); and TCF4 (NM_030756) (SEQ ID NO:43), or variants thereof having at least 80% nucleic acid sequence identity, wherein the tissue is from esophagus or colon, and wherein the expression in the test sample is at least 1.5-fold above baseline expression in a normal tissue control of the same tissue type.

7. A method according to claim 6, wherein the tissue is human tissue.

8. A method according to claim 6, wherein at least one of the at least eight genes is selected from the group consisting of AGR2 (SEQ ID NO:3), TM7SF1 (SEQ ID NO:13), MAT2B (SEQ ID NO:17), SLNAC1 (SEQ ID NO:23), and TCF4 (SEQ ID NO:43), or variants thereof
 5 having at least 80% nucleic acid sequence identity.

9. A method of detecting high-grade dysplasia (HGD) in cells of a mammalian tissue sample, the method comprising:

- 10 (a) obtaining a test tissue sample suspected of comprising cells exhibiting HGD;
- (b) establishing the level of expression in the test tissue sample of at least eight polypeptides encoded by genes selected from the group consisting of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3); ADAM8 (NM_001109) (SEQ ID NO:5); PRSS8 (Prostasin
 15 precursor, serine protease, NM_002773) (SEQ ID NO:7); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11); TM7SF1 (NM_003272) (SEQ ID NO:13); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19);
 20 PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID
 25 NO:31); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41); and TCF4 (NM_030756) (SEQ ID NO:43), or variants thereof
 30 having at least 80% nucleic acid sequence identity, wherein the tissue is from esophagus or colon; and
- (c) comparing expression of the at least eight polypeptides in the test tissue sample to expression of the at least eight polypeptides in normal tissue controls of the same tissue type, wherein an increase of at least 1.5-fold in expression of the polypeptides in the test tissue

sample relative to the normal tissue controls indicates that cells of the test sample exhibit HGD.

10. A method as according to claim 9 comprising contacting the test tissue sample with an antibody that specifically binds one of the at least eight polypeptides under conditions that permit the antibody to bind the polypeptide.

11. A method according to claim 9, wherein at least one of the at least eight polypeptides expressed by a gene selected from the group consisting of AGR2 (SEQ ID NO:3), TM7SF1 (SEQ ID NO:13), MAT2B (SEQ ID NO:17), SLNAC1 (SEQ ID NO:23), and TCF4 (SEQ ID NO:43), or variants thereof having at least 80% nucleic acid sequence identity.

12. The method of claim 1, wherein gene expression is determined by nucleic acid microarray analysis.

13. The method of claim 12, wherein analysis comprises contacting nucleic acid from a test tissue sample with a nucleic acid microarray comprising nucleic acid probe sequences, wherein at least eight of the nucleic acid probe sequences separately comprises at least 50 contiguous nucleotides from a gene selected from the group consisting of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3); ADAM8 (NM_001109) (SEQ ID NO:5); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11); TM7SF1 (NM_003272) (SEQ ID NO:13); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23); CAH4 (carbonic anhydrase iv precursor, NM_000717) (SEQ ID NO:25); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35); PHYH (phytanoyl-CoA-hydroxylase

(Refsum disease), NM_006214) (SEQ ID NO:37); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39); COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41); and TCF4 (NM_030756) (SEQ ID NO:43), or variants thereof having at least 80% nucleic acid sequence identity..

5

14. The method of claim 13, wherein the at least eight nucleic acid probe sequences comprise at least 60 contiguous nucleotides from a gene selected from the group.

15. The method of claim 14, wherein the at least eight nucleic acid probe sequences comprise
10 at least 80 contiguous nucleotides from a gene selected from the group.

16. The method of claim 15, wherein the at least eight nucleic acid probe sequences comprise at least 100 contiguous nucleotides from a gene selected from the group.

15 17. The method of claim 16, wherein the at least eight nucleic acid probe sequences comprise at least 150 contiguous nucleotides from a gene selected from the group.

18. The method of claim 17, wherein the at least eight nucleic acid probe sequences comprise at least 200 contiguous nucleotides from a gene selected from the group.

20

19. The method of claim 13, wherein the nucleic acid microarray comprises nucleic acid probe sequences from at least ten genes selected from the group.

20. The method of claim 19, wherein the nucleic acid microarray comprises nucleic acid
25 probe sequences from at least twelve genes selected from the group.

21. The method of claim 20, wherein the nucleic acid microarray comprises nucleic acid probe sequences from at least fifteen genes selected from the group.

30 22. The method of claim 21, wherein the nucleic acid microarray comprises nucleic acid probe sequences from at least eighteen genes selected from the group.

23. The method of claim 22, wherein the nucleic acid microarray comprises nucleic acid probe sequences from at least twenty genes selected from the group.

24. The method of claim 23, wherein the nucleic acid microarray comprises nucleic acid probe sequences from at least twenty two genes selected from the group.

25. The method of claim 1, wherein gene expression is determined by nucleic acid hybridization under high stringency conditions of a detectable probe comprising at least 50 contiguous nucleotides from a gene selected from the group to nucleic acid of cells of the test tissue sample relative to cells of the normal tissue control.

26. The method of claim 25, wherein the hybridization is *in situ* hybridization.

27. The method of claim 26, wherein the hybridization is fluorescent *in situ* hybridization.

28. The method of claim 1, wherein gene expression is determined by polymerase chain reaction (PCR) analysis.

29. The method of claim 1, wherein gene expression is determined by real-time polymerase chain reaction (RT-PCR) analysis.

30. The method of claim 1, wherein gene expression is determined by Taqman® polymerase chain reaction analysis.

31. A kit comprising a microarray, the microarray comprising nucleic acid probe sequences, wherein at least eight of the nucleic acid probe sequences each comprise at least 50 contiguous nucleotides from a gene selected from the group consisting of ET-1 (endothelin-1, NM_001955) (SEQ ID NO:1); AGR2 (anterior gradient 2 (*Xenopus laevis*) homolog, NM_006408) (SEQ ID NO:3); ADAM8 (NM_001109) (SEQ ID NO:5); PRSS8 (Prostasin precursor, serine protease, NM_002773) (SEQ ID NO:7); AXO1 (Axonin-1 precursor, NM_005076) (SEQ ID NO:9); NROB2 (Nuclear hormone receptor, NM_021969) (SEQ ID NO:11); TM7SF1 (NM_003272) (SEQ ID NO:13); DLDH (dihydrolipamide dehydrogenase, NM_000108) (SEQ ID NOS:15); MAT2B (methionine adenosyltransferase II, beta, NM_013283) (SEQ ID NO:17); STC-2 (stanniocalcin-2, NM_003714) (SEQ ID NO:19); PPBI (alkaline phosphatase, intestinal precursor, NM_001631) (SEQ ID NO:21); SLNAC1 (sodium channel receptor SLNAC1, NM_004769) (SEQ ID NO:23); CAH4 (carbonic

anhydrase iv precursor, NM_000717) (SEQ ID NO:25); PA21 (phospholipase a2 precursor, NM_000928) (SEQ ID NO:27); PAR2 (proteinase activated receptor 2 precursor, NM_005242) (SEQ ID NO:29); IDE (insulin-degrading enzyme, NM_004969) (SEQ ID NO:31); MYO1A (myosin-1A, NM_005379) (SEQ ID NO:33); CYP2J2 (cytochrome P450 monooxygenase, NM_000775) (SEQ ID NO:35); PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214) (SEQ ID NO:37); CYB5 (cytochrome b5, 3' end, NM_001914) (SEQ ID NO:39); COXVIb (cox VIb gene, last exon and flanking sequence, NM_001863) (SEQ ID NO:41); and TCF4 (NM_030756) (SEQ ID NO:43), or variants thereof having at least 80% nucleic acid sequence identity, and a package insert indicating that the microarray is for use in detecting HGD in a test tissue sample, wherein the tissue is from esophagus or colon, and wherein an increase in expression in the test tissue sample of at least 1.5-fold of the at least eight genes relative to a normal tissue control of the same tissue type indicates that cells of the test tissue exhibit HGD.

32. The kit of claim 31, wherein the nucleic acid probe sequences each comprise at least 60 contiguous nucleotides from a gene selected from the group.

33. The kit of claim 32, wherein the nucleic acid probe sequences each comprise at least 80 contiguous nucleotides from a gene selected from the group.

34. The kit of claim 33, wherein the nucleic acid probe sequences each comprise at least 100 contiguous nucleotides from a gene selected from the group.

35. The kit of claim 34, wherein the nucleic acid probe sequences each comprise at least 150 contiguous nucleotides from a gene selected from the group.

36. The kit of claim 35, wherein the nucleic acid probe sequences each comprise at least 200 contiguous nucleotides from a gene selected from the group.

37. A method of detecting cancer in a patient, the method comprising:

(a) obtaining a test tissue sample from the patient;

(b) establishing the level of expression of a gene selected from the group consisting of CAD17 (liver-intestine cadherin, NM_004063) (SEQ ID NO:45), CLDN15 (claudin 15, NM_014343) (SEQ ID NO:47), SLNAC1 (sodium channel, NM_004769) (SEQ ID NO:23),

CFTR (chloride channel, NM_000492) (SEQ ID NO:49), H2R (histamine H2 receptor, NM_022304) (SEQ ID NO:51), PRSS8 (serine protease, NM_002773) (SEQ ID NO:7), PA21 (phospholipase A2 group IB, NM_000928) (SEQ ID NO:27), AGR2 (anterior gradient 2 homolog, (NM_006408) (SEQ ID NO:3), EGFR (NM_005228) (SEQ ID NO:53), EPHB2 (NM_004442) (SEQ ID NO:55), CRIPTO CR-1 (NM_003212) (SEQ ID NO:57), Eprin B1 (NM_004429) (SEQ ID NO:59), MMP-17/MT4-MMP (NM_016155) (SEQ ID NO:61), MMP26 (NM_021801) (SEQ ID NO:63), ADAM10 (NM_001110) (SEQ ID NO:65), ADAM8 (NM_001109) (SEQ ID NO:5), ADAM1 (XM_132370) (SEQ ID NO:67), TIM1 (NM_003254) (SEQ ID NO:69), MUC1 (XM_053256) (SEQ ID NO:71), CEA (NM_004363) (SEQ ID NO:73), NCA (NM_002483) (SEQ ID NO:75), Follistatin (NM_006350) (SEQ ID NO:77), Claudin 1 (NM_021101) (SEQ ID NO:79), Claudin 14 (NM_012130) (SEQ ID NO:81), tenascin-R (NM_003285) (SEQ ID NO:83), CAD3 (NM_001793) (SEQ ID NO:85), AXO1 (NM_005076) (SEQ ID NO:9), CONT (NM_001843) (SEQ ID NO:87), Osteopontin (NM_000582) (SEQ ID NO:89), Galectin 8 (NM_006499) (SEQ ID NO:91), PGS1 (bihlycan, NM_001711) (SEQ ID NO:93), Frizzled 2 (NM_001466) (SEQ ID NO.:95), ISLR (NM_005545) (SEQ ID NO:97), FLJ23399 (NM_022763) (SEQ ID NO:99), TEM1 (NM_020404) (SEQ ID NO:101), Tie2 ligand2 (NM_001147) (SEQ ID NO:103), STC-2 (NM_003714) (SEQ ID NO:19), VEGFC (NM_005429) (SEQ ID NO:105), tPA (NM_000930) (SEQ ID NO:107), Endothelin 1 (NM_001955) (SEQ ID NO:1), Thrombomodulin (NM_000361) (SEQ ID NO:109), TF (NM_001993) (SEQ ID NO:111), GPR4 (NM_005282) (SEQ ID NO:113), GPR66 (NM_006056) (SEQ ID NO:115), SLC22A2 (NM_003058) ((SEQ ID NO:117), MLSN1 (NM_002420) (SEQ ID NO:119), and ATN2 (Na/K transport, NM_000702) (SEQ ID NO:121), or variants thereof having at least 80% nucleic acid sequence identity, wherein the test tissue is from esophagus or colon; and wherein the expressing in the test tissue is at a level at least 1.5-fold above expression of the gene in a normal tissue control of the same tissue type.

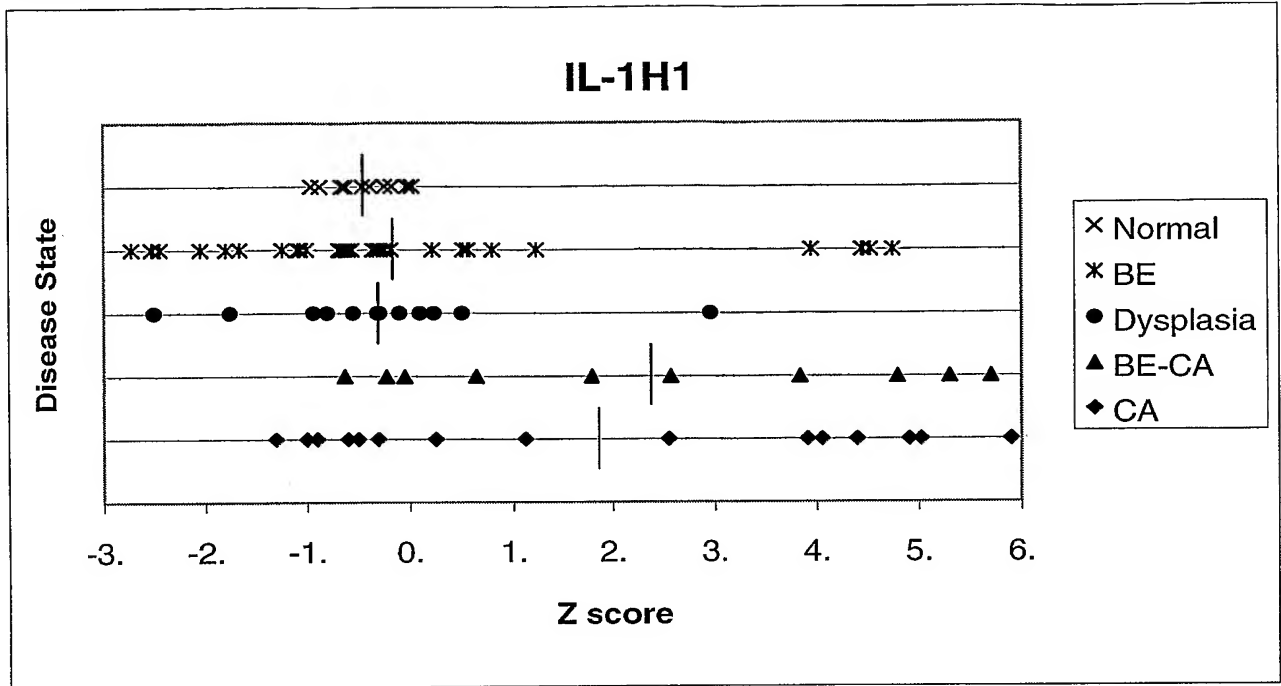
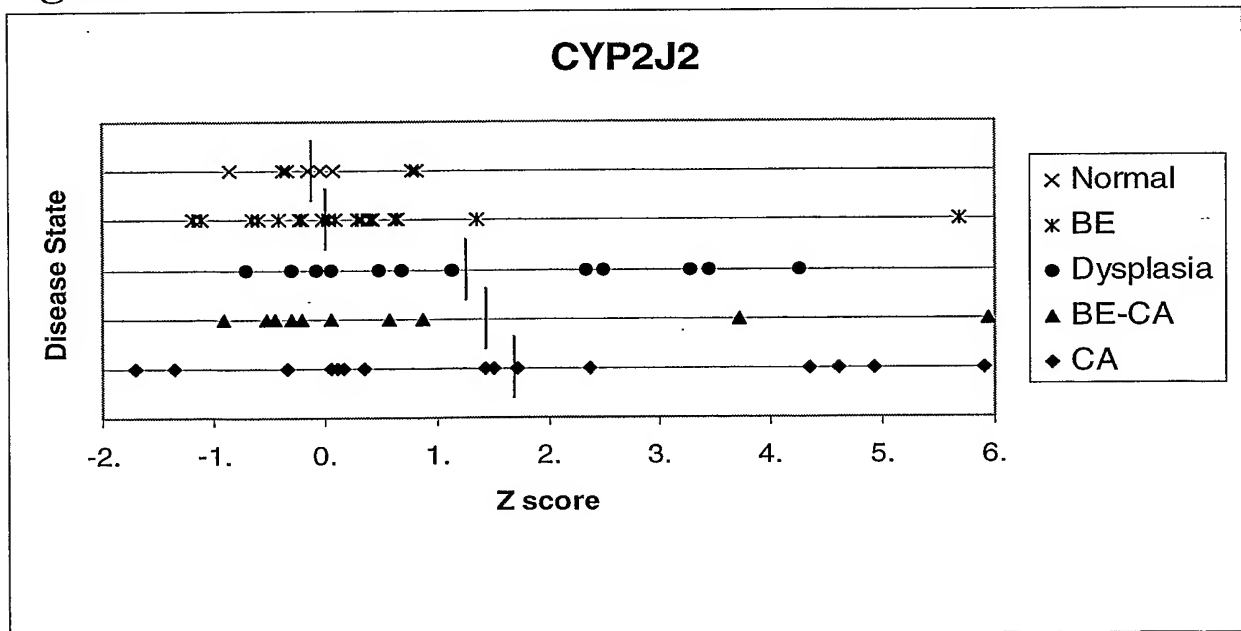
38. The method of claim 37, wherein inhibition of cell growth is cell death.

39. The method of claim 37, wherein at least two genes selected from the group are expressed at a level at least 1.5-fold above expression of the gene in a normal cell control.

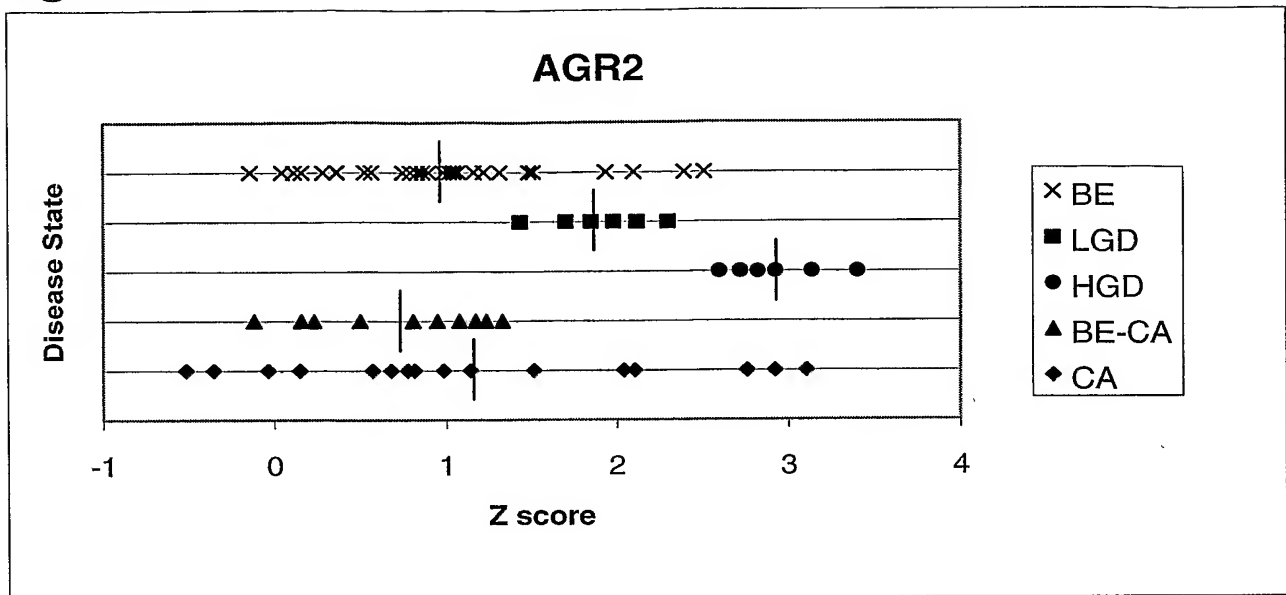
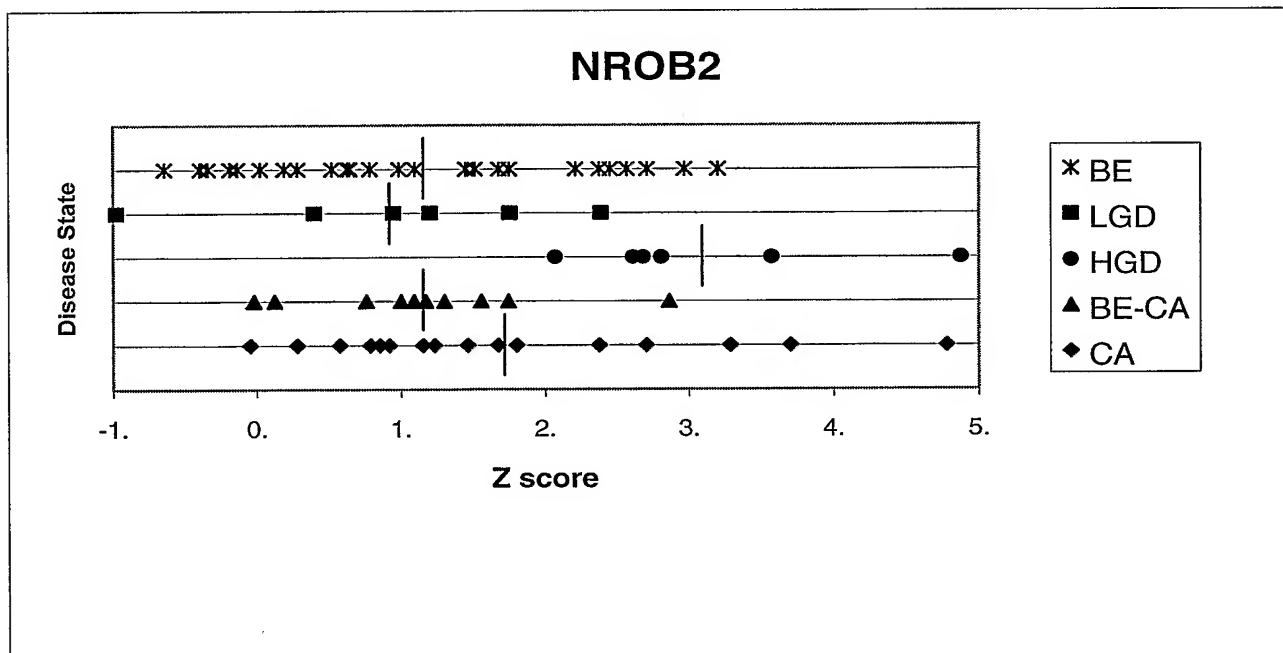
40. The method of claim 39, wherein at least three genes selected from the group are expressed at a level at least 1.5-fold above expression of the gene in a normal cell control.

41. The method of claim 40, wherein at least 5 genes selected from the group are expressed at a level at least 1.5-fold above expression of the gene in a normal cell control.
- 5 42. The method of claim 41, wherein at least 8 genes selected from the group are expressed at a level at least 1.5-fold above expression of the gene in a normal cell control.
43. The method of claim 1, wherein the expression p value is less than 0.07.
- 10 44. The method of claim 6, wherein the expression p value is less than 0.07.
45. The method of claim 9, wherein the expression p value is less than 0.07.

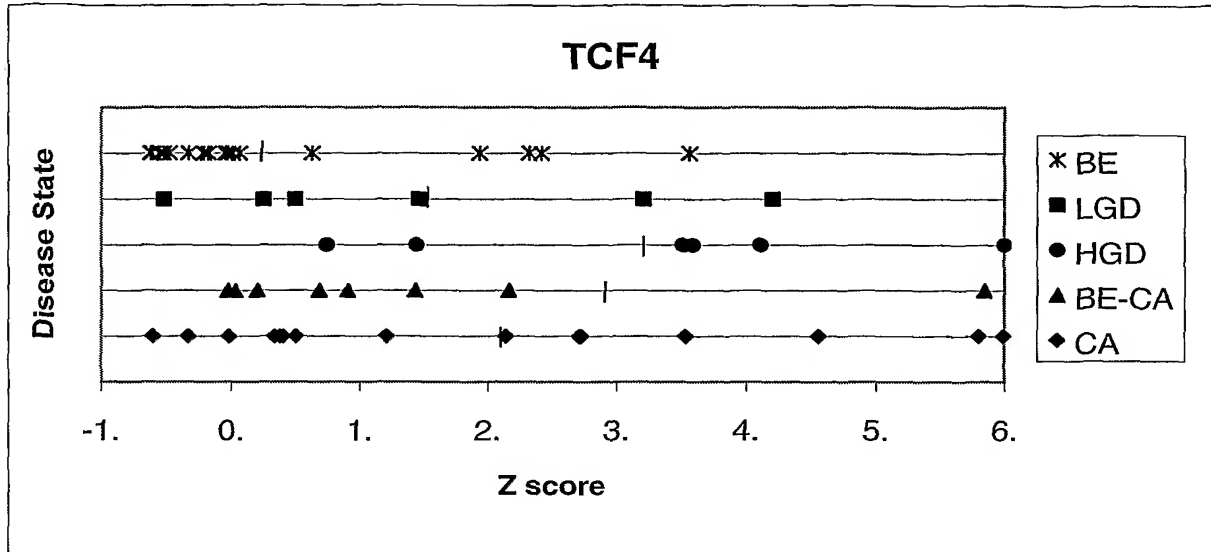
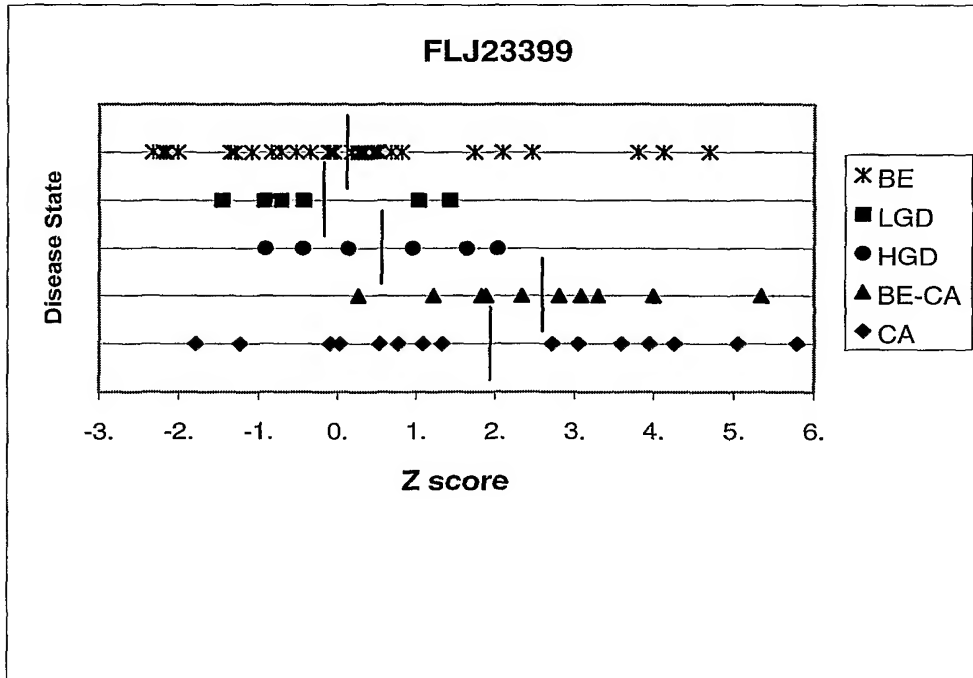
1/115

Figure 1A**Figure 1B**

2/115

Figure 2A**Figure 2B**

3/115

Figure 3A**Figure 3B**

4/115

ET-1 (endothelin-1, NM_001955)

```

1  cgccgcgtgc gcctgcagac gctccgctcg ctgccttctc tcctggcagg cgctgccttt
61  tctccccgtt aaagggcact tgggctgaag gatcgctttg agatctgagg aaccgcgagc
121 gcttttgagg acctgaagct gtttttcttc gttttccttt gggttcagtt tgaacgggag
181 gtttttgatc ctttttttcc agaatggatt atttgctcat gattttctct ctgctggttg
241 tggcttgcca aggagctcca gaaacagcag tcttaggcgc tgagctcagc gcggtgggtg
301 agaacggcgg ggagaaaccc actcccagtc caccctggcg gctccgccgg tccaagcgct
361 gtccttgctc gtccctgatg gataaagagt gtgtctactt ctgccacctg gacatcattt
421 gggccaacac tcccagcac gttgttccgt atggacttgg aagccctagg tccaagagag
481 ccttggagaa ttacttccc acaaaggcaa cagaccgtga gaatagatgc caatgtgcta
541 gccaaaaaga caagaagtgc tgggaattttt gccaaagcagg aaaagaactc agggctgaag
601 acattatgga gaaagactgg aataatcata agaaaggaaa agactgttcc aagcttggga
661 aaaagtgtat ttatcagcag ttagtgagag gaagaaaaat cagaagaagt tcagaggaac
721 acctaaagaca aaccaggtcg gagaccatga gaaacagcgt caaatcatct tttcatgatc
781 ccaagctgaa aggcaatccc tccagagagc gttatgtgac ccacaaccga gcacattggt
841 gacagacctt cggggcctgt ctgaagccat agcctccacg gagagccctg tggccgactc
901 tgcactctcc accctggctg ggatcagagc aggagcatcc tctgctgggt cctgactggc
961 aaaggaccag cgtcctcggt caaaacattc caagaaaggt taaggagttc ccccaaccat
1021 cttcactggc ttccatcagt ggtaactgct ttggtctctt ctttcatctg gggatgacaa
1081 tggacctctc agcagaaaca cacagtcaca ttcgaattcg ggtggcatcc tccggagaga
1141 gagagaggaa ggagattcca cacaggggtg gagtttctga cgaaggctct aagggagtgt
1201 ttgtgtctga ctcaggcgcc tggcacattt caggggagaaa ctccaaagtc cacacaaaga
1261 ttttctaagg aatgcacaaa ttgaaaacac actcaaaaga caaacatgca agtaaagaaa
1321 aaaaaaaaaa aaaa (SEQ ID NO:1)

```

FIGURE 4A

ET-1 (endothelin-1, NM_001955)

```

MDYLLMIFSLLFVACQGAPETAVLGAELSAVGENGGEKPTSPSP
RLRRSKRCSCSSLMDKECVYFCHLDIIWVNTPEHVVPYGLGSPRSKRALENLLPTKA
TDRENRCQCASQKDKKWNFCQAGKELRAEDIMEKDWNHKKGKDCSKLGKKCIYQQL
VRGRKIRRSSEEHLRQTRSETMRNSVKSSFHDPKLGKGNPSRERYVTHNRAHW (SEQ ID NO:2)

```

FIGURE 4B

5/115

AGR2 (anterior gradient 2 (Xenopus laevis) homolog, NM_006408)

```

1  ccgcaccta gccgccgact cacacaaggc aggtgggtga ggaaatccag agttgccatg
61 gagaaaattc cagtgtcagc attcttgctc cttgtggccc tctoctacac tctggccaga
121 gataccacag tcaaacctgg agccaaaaag gacacaaagg actctcgacc caaactgccc
181 cagaccctct ccagaggttg ggggtgaccaa ctcatctgga ctacagacata tgaagaagct
241 ctatataaat ccaagacaag caacaaaccc ttgatgatta ttcatcactt ggatgagtgc
301 ccacacagtc aagcttttaa gaaagtgttt gctgaaaata aagaaatcca gaaattggca
361 gagcagtttg tcctcctcaa tctggtttat gaaacaactg acaaacacct ttctcctgat
421 ggcagtatg tcccaggat tatgtttgtt gacccatctc tgacagttag agccgatatc
481 actggaagat attcaaactg tctctatgct tacgaacctg cagatacagc tctgttgctt
541 gacaacatga agaaagctct caagttgctg aagactgaat tgtaaagaaa aaaaatctcc
601 aagcccttct gtctgtcagg ccttgagact tgaaaccaga agaagtgtga gaagactggc
661 tagtgtggaa gcatagtga cactgtgatt aggttatggg ttaatgttac aacaactatt
721 ttttaagaaa aacaagtttt agaaatttgg tttcaagtgt acatgtgtga aaacaatatt
781 gtatactacc atagtgaacc atgattttct aaaaaaaaaa ataatgttt tgggggtgtt
841 ctgttttctc caacttggtc tttcacagtg gttcgtttac caaataggat taaacacaca
901 caaatgctc aaggaagggg caagacaaaa ccaaaactag ttcaaagat gaagacaaa
961 gaccaagtta tcatctcacc acaccacagg ttctcactag atgactgtaa gtagacacga
1021 gcttaatcaa cagaagtatc aagccatgtg ctttagcata aaagaatatt tagaaaaaca
1081 tcccaagaaa atcacatcac tacctagagt caactctggc caggaactct aaggtaacac
1141 ctttcattta gtaattaaat tttagtcaga ttttgcccaa cctaagctc tcagggaaag
1201 cctctggcaa gtagctttct ccttcagagg tctaatttag tagaaaggct atccaaagaa
1261 catctgcact cctgaacaca cctgaagaa atcctgggaa ttgaccttgt aatcgatttg
1321 tctgtcaagg tcctaaagta ctggagtga ataaattcag ccaacatgtg actaattgga
1381 agaagagcaa aggggtgtga cgtgttgatg aggcagatgg agatcagagg ttactagggt
1441 ttaggaaacg tgaaaggctg tggcatcagg gtaggggagc attctgcta acagaaatta
1501 gaattgtgtg ttaatgtctt cactctatac ttaatctcac attcattaat atatggaatt
1561 cctctactgc ccagccctc ctgatttctt tggccctgg actatggtgc tgtatataat
1621 gctttgcagt atctgttgct tgtcttgatt aacttttttg gataaaacct tttttgaaca
1681 gaaaaaaaaa aaaaaaaaaa a (SEQ ID NO:3)

```

FIGURE 5A

AGR2 (anterior gradient 2 (Xenopus laevis) homolog, NM_006408)

```

MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTL
SRGWGDQLIWTQTYEEALYKSKTSNKPLMI IHHLDECPHSQALKKVFAENKEIQKLAE
QFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALL
LDNMKKALKLLKTEL (SEQ ID NO:4)

```

FIGURE 5B

6/115

ADAM8 (NM_001109)

```

1  gacccgggcca tgcgcggcct cgggctcttg ctgctgggcg cgatgatgct gcctgcgatt
61  gccccagacc ggccctgggc cctcatggag cagtatgagg tcgtgttgcc gggcgctctg
121 ccaggccccc gagtccgccc agctctgccc tccacttggt gcctgcaccc agagaggggtg
181 agctacgtcc ttggggccac agggcacaac ttcaccctcc acctgcggaa gaacaggggac
241 ctgctgggtt ccggctacac agagacctat acggctgcca atggctccga ggtgacggag
301 cagcctcgcg ggcaggacca ctgcttatac cagggccacg tagaggggta cccggactca
361 gccgccagcc tcagcacctg tgccggcctc aggggtttct tccaggtggg gtacagacctg
421 cacctgatcg agcccctgga tgaaggtggc gagggcggac ggcacgccgt gtaccaggct
481 gagcacctgc tgcagacggc cgggacctgc ggggtcagcg acgacagcct gggcagcctc
541 ctggggacccc ggacggcagc cgtcttcagg cctcgccccg gggactctct gccatcccga
601 gagacccgct acgtggagct gtatgtggtc gtggacaatg cagagttcca gatgctgggg
661 agcgaagcag ccgtgcgtca tcgggtgctg gaggtggtga atcacgtgga caagctatat
721 cagaaactca acttccgtgt ggtcctggtg ggcttggaag tttggaatag tcaggacagg
781 ttccacgtca gccccgaccc cagtgtcaca ctggagaacc tcctgacctg gcaggcacgg
841 caacggacac ggccggcacct gcatgacaac gtacagctca tcacgggtgt cgacttcacc
901 gggactactg tggggtttgc cagggtgtcc gccatgtgct cccacagctc aggggctgtg
961 aaccaggacc acagcaagaa ccccgctggc gtggcctgca ccatggccca tgagatgggc
1021 cacaacctgg gcatggacca tgatgagaac gtccagggtc gccgctgcca ggaacgcttc
1081 gaggccggcc gctgcatcat ggcaggcagc attggctcca gtttccccag gatgttcagt
1141 gactgcagcc aggcctacct ggagagcttt ttggagcggc cgcagtcggt gtgcctcgcc
1201 aacgcccctg acctcagcca cctggtgggc ggccccgtgt gtgggaacct gtttgtggag
1261 cgtggggagc agtgcgactg cggccccccc gaggactgcc ggaaccgctg ctgcaactct
1321 accacctgcc agctggctga gggggccccg tgtgcgcacg gtacctgctg ccaggagtgc
1381 aaggtgaagc cggctggtga gctgtgccgt cccaagaagg acatgtgtga cctcgaggag
1441 ttctgtgacg gccggcaccc tgagtgcctc gaagacgcct tccaggagaa cggcacgccc
1501 tgctccgggg gctactgcta caacggggcc tgtccacac tggcccagca gtgccaggcc
1561 ttctgggggc caggtgggca ggctgccgag gagtctgtct tctcctatga catcctacca
1621 ggctgcaagg ccagccggta cagggtgac atgtgtggcg ttctgcagtg caaggggtggg
1681 cagcagcccc tggggcgtgc catctgcac gtggatgtgt gccacgcgct caccacagag
1741 gatggcactg cgtatgaacc agtgcccagc ggcacccggt gtggaccaga gaaggtttgc
1801 tggaaaggac gttgccagga cttacacggt tacagatcca gcaactgctc tgcccagtgc
1861 cacaacctat ggggtgtgaa ccacaagcag gagtgccact gccacgcggg ctgggccccg
1921 cccactgctg cgaagctgct gactgaggtg cacgcagcgt ccgggagcct cccgctctc
1981 gtggtggtgg ttctggtgct cctggcagtt gtgctggtca cctggcagg catcatcgtc
2041 taccgcaaa cccggagccg catcctgagc aggaacgtgg ctcccagac cacaatgggg
2101 cgctccaacc ccctgttcca ccaggtgcc agccgcgtgc cggccaaggg cggggctcca
2161 gccccatcca ggggccccca agagctggtc cccaccacc acccgggcca gcccgccga
2221 caccggcct cctcggtggc tctgaagagg ccgccccctg ctctccggt cactgtgtcc
2281 agcccaccct tcccagttcc tgtctacacc cggcaggcac caaagcaggt catcaagcca
2341 acgttcgcac cccagtgcc cccagtcaaa cccggggctg gtgcggccaa ccctggtcca
2401 gctgaggggt ctggtggccc aaaggttgcc ctgaagcccc ccatccagag gaagcaagga
2461 gccggagctc ccacagcacc ctaggggggc acctgcgct gtgtggaaat ttggagaagt
2521 tgcggcagag aagccatgct ttccagcctt ccacgggtcca gctagtgcg ctacgcctta
2581 gacctgact ttgcaggct agctgctggt ctaacctcag taatgcatct acctgagagg
2641 ctctgctgt ccacgcctc agccaattcc ttctccccgc cttggccacg tgtagcccca
2701 gctgtctgca ggcaccaggc tgggatgagc tgtgtgcttg cgggtgcgtg tgtgtgtacg
2761 tgtctccagg tggccgctgg tctcccgctg tggtcaggag gccacatata cagcccctcc
2821 cagccacacc tgcccctgct ctggggcctg ctgagccggc tgcccggggc acccggttcc
2881 aggcagcaca gacgtggggc atcccagaa agactccatc ccaggaccag gttcccctcc
2941 gtgctcttcg agaggggtgt agtgagcaga ctgcaccca agctcccgac tccaggtccc
3001 ctgatcttgg gcctgtttcc catgggattc aagagggaca gcccagctt tgtgtgtgtt
3061 taagcttagg aatgcctttt atggaaaggg ctatgtggga gactagcta tcttgtctgg
3121 tttctttgag acctcagatg tgtgttcagc agggctgaaa gcttttatto ttttaataatg
3181 agaaatgtat attttactaa taaattattg accgagttct gtagattctt gttaga (SEQ

```

ID NO:5)

FIGURE 6A

7/115

ADAM8 (NM_001109)

MRGLGLWLLGAMMLPAIAPSRPWALMEQYEVVLPRRLPGPRVRR
ALPSHLGLHPERVSYVLGATGHNFTLHLRKNRDLGSGYTETTYTAANGSEVTEQPRGQ
DHCLYQGHVEGYPDASAASLSTCAGLRGFFQVGSDLHLIEPLDEGGEGGRHAVYQAEHL
LQTAGTCGVSDDSLGSLLGPRTAAVFRPRPGDSLPSRETRYVELYVVVDNAEFQMLGS
EAAVRHRVLEVVNHVDKLYQKLNFRVVLVGLLEIWNQDRFHVSPDPSVTLENLLTWQA
RQRTRRHLHDNVQLITGVDFGTGTTVG FARVSAMCSHSSGAVNQDHSKNPVG VACTMAH
EMGHNLGMDHDENVQGCRCQERFEAGRCIMAGSIGSSFP RMFSDCSQAYLESFLERPQ
SVCLANAPDLSHLVGGPVCNLFVERGEQCDCGPPEDCRNRCCNSTTCQLAEGAQCAH
GTCCQECKVKPAGELCRPKKDMCDLEEFCDGRHPECPEDAFQENGTPCSGGYCYNGAC
PTLAQQCQAFWGP GPGQAAEESCF SYDILPGCKASRYRADMCGVLQCKGGQQPLGRAIC
IVDVCHALT TEDGTAYEPVPEGTRCGPEKVCWKGRQC DLHVYRSSNCSAQCHNHGVCN
HKQECHCHAGWAPPHCAKLLTEVHAASGSLPVLVVVVLVLLAVVLVTLAGIIVYRKAR
SRILSRNVAPKTTMGRSNPLFHQAASRVPAKGGAPAPSRGPQELVPTTHPGQPARHPA
SSVALKRPPPPAPPVTVSSPPFPVPVYTRQAPKQVIKPTFAPPVPPVKPGAGAAPNGPA
EGAVGPKVALKPPIQRKQGAGAPTAP (SEQ ID NO:6)

FIGURE 6B

8/115

PRSS8 (Prostasin precursor, serine protease, NM_002773)

```

1  gacttttggtg gcaagaggag ctggcgggagc ccagccagtg ggccggggcca ggggaggggc
61  gggcaggtag gtgcagccac tcctgggagg accctgcgtg gccagacggt gctggtgact
121 cgtccacact gctcgcttcg gatactccag gcgtctcccg ttgcggccgc tccttgcctt
181 agaggccagc cttggacact tgctgccccct ttccagcccg gattctggga tccttccctc
241 tgagccaaca tctgggtcct gccttcgaca ccacccaag gcttcctacc ttgctgcct
301 ggagtctgcc ccagggggcc ttgtcctggg ccatggccca gaagggggtc ctggggcctg
361 ggcagctggg ggctgtggcc attctgctct atcttggtt actccggtcg gggacaggag
421 cggaaggggc agaagctccc tgcggtgtgg cccccaagc acgcatcaca ggtggcagca
481 gtgcagtcgc cggtcagtgg ccctggcagg tcagcatcac ctatgaaggc gtccatgtgt
541 gtggtggctc tctcgtgtct gagcagtggg tgctgtcagc tgctcactgc tccccagcg
601 agcaccacaa ggaagcctat gaggtcaagc tggggggcca ccagctagac tcctactccg
661 aggacgccaa ggtcagcacc ctgaaggaca tcatcccca cccagctac ctccaggagg
721 gctcccaggg cgacattgca ctccctccaac tcagcagacc catcaccttc tcccgctaca
781 tccggcccat ctgcctccct gcagccaacg cctccttccc caacggcctc cactgcactg
841 tcactggctg gggtcagtgt gccccctcag tgagcctcct gacgccaag ccactgcagc
901 aactcgaggt gcctctgac agtcgtgaga cgtgtaactg cctgtacaac atcgacgcca
961 agcctgagga gccgcacttt gtccaagagg acatgggtgt tgctggctat gttggagggg
1021 gcaaggacgc ctgccagggt gactctgggg gccactctc ctgccctgtg gagggtctct
1081 ggtacctgac gggcatttgt agctggggag atgctgtgg ggcccgaac aggcctggtg
1141 tgtacactct ggcctccagc tatgcctcct ggatccaaag caaggtgaca gaactccagc
1201 ctcgtgtggt gcccctaaac caggagtccc agcccagacg caacctctgt ggcagccacc
1261 tggccttcag ctctgcccc acccagggtc tgctgaggcc catccttttc ctgcctctgg
1321 gcctggctct gggcctcctc tccccatggc tcagcgagca ctgagctggc cctacttcca
1381 ggatggatgc atcacactca aggacaggag cctggtcctt ccctgatggc ctttggacct
1441 agggcctgac ttgagccact ccttccttca ggactctgcg ggaggctggg gccccatctt
1501 gatctttgag cccattcttc tgggtgtgct ttttgggacc atcactgaga gtcaggagtt
1561 ttactgcctg tagcaatggc cagagcctct ggccctcac ccacctatga ccagccatt
1621 ggccgagctc ctggggagct cctgggacct ttggctatga aaatgagccc tggctccac
1681 ctgtttctgg aagactgctc ccggcccgc tgcccagact gatgagcaca tctctctgcc
1741 ctctccctgt gttctgggct ggggccacct ttgtgcagct tcgaggacag gaaaggcccc
1801 aatcttgccc actggccgct gaggccccc gagccctgac tcctggactc cggaggactg
1861 agccccacc ggaactgggc tggcgcttgg atctggggtg ggagtaacag ggcagaaatg
1921 attaaaatgt ttgagcac (SEQ ID NO:7)

```

Figure 7A

9/115

PRSS8 (Prostasin precursor, serine protease, NM_002773)

MAQKGVLGPGQLGAVAILLYLGLLRSGTGAEGAEAPCGVAPQAR
ITGGSSAVAGQWPWQVSITYEGVHVCGGSLVSEQWVLSAAHCFPSEHHKEAYEVKLGA
QLDSYSEDAKVSTLKDIIHPHPSYLQEGSQGDIALQLSRPITFSRYIRPICLPAANA
SFPNGLHCTVTGWGHVAPSVSLLTPKPLQQLEVPLISRETCNCLYNIDAKPEEPHFVQ
EDMVCAGYVEGGKDACQGDGGPLSCPVEGLWYLTGIVSWG DACGARNRPGVYTLASS
YASWIQSKVTELQPRVVPQTQESQPD SNLCGSHLAFSSAPAQGLLRPILFLPLGLALG
LLSPWLSEH (SEQ ID NO:8)

Figure 7B

10/115

AXO1 (Axonin-1 precursor, NM_005076)

```

1  acacacacgc gccctcaccc gccaccgccg ccgcggccgc cgccgcaccc ggacagcgag
61  cggctgaggg cgccaggggc caaaggacag cggcccagac aggggctggc ggcccggccg
121  gccccggctc accgactcgg gcagcatcca cctgccccag ccaacaccct tctctcgccc
181  caggtccttt ctcagcctcc agctgggctg tccccaaagt gagctgaggg tcttctcctc
241  cgatccccac ctctgcccgg acatccacca tggggacagc caccaggagg aagccacacc
301  tgctgctggt agctgctgtg gcccttgtct cctcttcagc ttggagttca gccctgggat
361  cccaaaccac cttcgggcct gtctttgaag accagccctt cagtgtgcta tccccagagg
421  agtccacgga ggagcagggt ttgctggcat gccgcgcccg ggccagccct ccagccacct
481  atcggtgga gatgaatggt accgagatga agctggagcc aggttcccg t caccagctgg
541  tggggggcaa cctggtcatc atgaaccca ccaaggcaca g gatgccggg gtctaccagt
601  gcctggcctc caaccagtg ggcaccgttg gcagcaggga ggccatcctc cgcttcggct
661  ttctgcagga attctccaag gaggagcgag acccagtgaa agctcatgaa ggctgggggg
721  tgatgttgcc ctgtaaccca cctgccactt acccaggctt gtctaccgc tggctcctca
781  acgagttccc caacttcac cgcagcgagc ggcgtcactt cgtgtccag accacaggga
841  acctgtacat tgcccgaacc aatgcctcag acctgggcaa ctactcctgt ttggccacca
901  gccacatgga cttctccacc aagagcgtct tcagcaagtt tgctcagctc aacctggctg
961  ctgaagatac ccggctcttt gcaccagca tcaaggcccg gtcccagca gagacctatg
1021  cactgggtgg gcagcagggt accctggagt gcttcgcctt tgggaacctt gtcccccgga
1081  tcaagtggcg caaagtggac ggctcctgtt ccccgagtg gaccacagt gacccacc
1141  tgcagatccc cagcgtcagc tttgaggatg agggcaccta cgagtgtgag cgggagaact
1201  ccaagggccg agacaccgtg cagggccgca tcatcgtgca ggctcagcct gagtggctaa
1261  aagtgatctc ggacacagag gctgacattg gctccaaact gcgttggggc tgtgcagccg
1321  ccggcaagcc ccggcctaca gtgcgttggc tgcggaacgg ggagcctctg gcctcccaga
1381  accgggtgga ggtgttggct ggggacctgc ggttctccaa gctgagcctg gaagactcgg
1441  gcatgtacca gtgtgtggca gagaataagc acggtaccat ctacgccagc gccgagctag
1501  ccgtgcaagc actcgccctt gacttcaggg tgaatcccg gaggcgtctg atcccccg
1561  cccgcggggg agagatcctt atccccctgc agccccgggc agtccaaag gccgtggtgc
1621  tctggagcaa aggcacggag attttgggtc acagcagcag agtgactgta actccagatg
1681  gcaccttgat cataaagaaac atcagccggt cagatgaagg caaatacacc tgccttgcctg
1741  agaacttcat gggcaaaagg aacagcactg gaatcctatc tgtgcgagat gcaacaaaaa
1801  tcaactctagc cccctcaagt gccgacatca acttgggtga caacctgacc ctacagtgcc
1861  atgcctccca cgacccacc atggacctca ccttcacctg gacctggac gacttcccca
1921  tcgactttga taagcctgga gggcactacc ggagaactaa tgtgaaggag accattgggg
1981  atctgaccat cctgaacgcc cagctgcgcc atggggggaa gtacacgtgc atggcccaga
2041  cggtggtgga cagcgcgtcc aaggaggcca cagtcctggt ccgaggtccg ccaggtcccc
2101  caggaggtgt ggtggtgagg gacattggcg acaccacctt ccagctcagc tggagccgtg
2161  gcttcgacaa ccacagcccc atcgctaagt acaccctgca agctcgact ccacctgcag
2221  ggaagtggaa gcaggttcgg accaatcctg caaacatcga gggcaatgcc gagactgcac
2281  aggtgctggg cctcaccccc tggatggact atgagttccg ggtcatagcc agcaacattc
2341  tgggcactgg ggagcctagt gggccctcca gcaaaatccg gaccaggga gacagccccct
2401  cggtgccacc ctcaggactc agcggaggag gtggagcccc cggagagctc atcgtcaact
2461  ggacgcccac gtcacgggag taccagaacg gagacggctt cggctacctg ctgtccttcc
2521  gcaggcaggg cagcactcac tggcagaccg cccgggtgcc tggcgccgat gccaggtact
2581  ttgtctacag caacgagagc gtccggccct acacgccctt tgaggtcaag atccgcagct
2641  acaaccgccg cggggatggg cccgagagcc tcactgcact cgtgtactca gctgaggaag
2701  agcccagggg gggccctacc aaggtgtggg ccaaaggggt ctatcctca gagatgaacg
2761  tgacctggga acccgtgcag caggacatga atggtatcct cctggggtat gagatccgct
2821  actggaaaagc tggggacaaa gaagcagctg cggaccgagt gaggacagca gggctggaca
2881  ccagtgcctg agtcagcggc ctgcaccca acaccaagta ccatgtgacc gtgagggcct
2941  acaaccgggc tggcactggg cctgccagcc cttctgcaa cgccacgacc atgaagcccc
3001  ctccgcggcg acctcctggc aacatctcct ggactttctc aagctctagt cttagcatta
3061  agtgggaccc tgtggtccct ttccgaaatg agtctgcagt caccggctat aagatgctgt

```

FIGURE 8A

11/115

```

3121 accagaatga cttacacctg actcccacgc tccacctcac cggcaagaac tggatagaaa
3181 tcccagtgcc tgaagacatt ggccatgccc tggtaaaaat tcggaccaca gggcccggag
3241 gggatgggat ccctgcagaa gtccacatcg tgaggaatgg aggcacaagc atgatggtgg
3301 agaacatggc agtccgcccc gcaccacacc ctggcaccgt catttccac tccgtggcga
3361 tgctgatcct cataggctcc ctggagctct gatcctggaa cccctccctc tgcgccgacg
3421 ctggacgcca cctccgacgg acacagccag ccccttcctg ctgccaaggt ggcccgacac
3481 tgtgccagag agtggctggg tttaaatacc tactttaaac agtgcccttt ttgtaggagg
3541 taggatattt tatattctgc cgcaggatag aaccacgca aggattttct ttaaattgag
3601 aggcaccagg cagtaacttc catgatgaca ctgacgccta tacctgagct ctaggctgcc
3661 tggaggggag gaacaggccc atgggaagaa gggggtttta aaaacatgtc tccaactcag
3721 cagagatggc cctctgggac cctatacgga ctccgccact tgagagcagt cctaggcccg
3781 gcaggaacac cagacatgaa caggttgaag aactggagcg aagtgcacac ctaccatcc
3841 ttcagtctaa ggaagaaggg caagccctgg gaccaagagc tctccgcct tctccctcga
3901 gcagcagcaa ggacctgac gctgtccccg ataactccct aggggctcct gcctgccaa
3961 ggggctgaga accagcgccc cgatgcctga ggctgggagc ctgagccctc tcagctttga
4021 ggggggtgat actccaggct gtttggggtg ggagccaaaa agagttgaga ggccaggggc
4081 cttggtggaa aggggcacca gccttggtct gagatagtca caaccagggt gacgatgccc
4141 tctcagccaa cactgccaac ctgaccctgt catcccgatt gacagcgcca cttcaggtgg
4201 ctgggtgact aaagggtctg tcttggtggg gtctcccacc cctccaagac ccattctgca
4261 cagtccctcc agggtttggg caggagatgg ccaatcatgc gccacacctc ccagtgtctg
4321 ctgcagtcat ctgggcctcc ccgacctgca gcccagact ctgctctccc agcactgact
4381 cactcctgcc tgggagggga atgcagcatt catgctgtgt gtctggtat tgggaggtt
4441 ctgggaaggg cagaggataa atgtggccct gcctgctccc aggtatacct aggaccacct
4501 ggccagatcc gctcccagac ggccctggac tgcttgcat tccccggaga aaaagggtt
4561 aataaatggg ccataccttc ctgagctctg ggtatactac cagtcacaga acgtcagagc
4621 tggagaagc cttagagctc aacttcttca agccctcac tttacagatg aggaaatgga
4681 ggtggtccag agagggtctg ggattcccaa ggtcacacag ccagaagag atggggctgg
4741 gttaagaact cgagtcttcc acctttctgt tcaaggctgt ttgtctaccc agaggaagga
4801 ggcactgctg aatggctatg gcctggctaa gaagggtgat agtcagtagg gtgtgaaaat
4861 tctacttcaa ggggttcgga ttggtgatca tggggattgg catggtctggg ttcccgcca
4921 aggtgtgggc agagcttcta ccaaacttca acatggaggg ctgacttgaa gctccctgtc
4981 cccctcactc ttgccccag aaaagaggcc aaagcaagag cagattccct agcaagagc
5041 agcagcacia ctaggaaacc ccaaagccca tgctccgaca ggtggccctt cacagggggc
5101 agcgggacag gcatcttgaa gggcatatgt cctcggaagc tccgagcctg ttttctgtag
5161 tttatagtta gagctctatt ttggttatgg tttttaact tttagtcct gctctatttt
5221 cctgggcagg tttatgttga tgtttacca ctacaatttt ttaaaaatat aagctcacat
5281 gccttttccc tgccacagcc aaacccccac tgcaccctac ccacccaccc ctaccccagg
5341 tcagctttcc tggagctggc taatgaaagc ctctcacct ctcccaacc cttacaagca
5401 agggctgctg gggctcagct atacgaccat tctccctgac agggagtcca aacttggcct
5461 agcatccctc ctggcccccc tctggccacg acttggcctg tgccctggtc tctatcagaa
5521 aggggatgct gaacaaaacc tccttccaag ttttatccaa ttctgtctcc attgcctcgg
5581 gctgcgtcag gggaagcagg ggacagggtg ccagttgctg ggccgaggga ggagctggtt
5641 tggcatagga cctaaccagt gaagctagag gctacagcca ctaaacttgc ttcaggccaa
5701 cgatagttac tcacaagtaa gtaccttaat gctaagtagg tccactaaaa aggggaggaa
5761 ggcagacctc ctgggagacc cacgaagggt ttttagccag ggaaaactga gccccaggaa
5821 aacctaacca ctgggcaggc agaatttgtt tgagggatag aacgacaaca aaataaatgt
5881 tcctgcagcc tgagatttca ggtagagtac tgactaagg ttaataagac aatagggtgac
5941 ctgaggacat gcaagcttgt aaaatgcaac agcctcctgc tagagtgact tgtacatgag
6001 cttgcttgca gaagactaga ttatagtttt ctcaggatcc cctcctgcgc aggggttctc
6061 tgattttcgt gttctctgcc cagatgggtc gggggagtgt agagtgtgct tattttcact
6121 gcgatcatga gaccacagtt ctgggttatc tcctctcata catcaagccc cagaggaggc
6181 ggcaagagga acagccacaa acaagtactt taccacacag cttagtggcc agtaaacacc

```

FIGURE 8B

12/115

```

6241 ctgggggacta ggaaaaggaa ccaactgtag gcacctctcc agggcctagg gagacaagtg
6301 tcctctcttc tgcatacatt tgggctcccc ttacagagcc ctttgccctg gctctctggg
6361 ccttggtgct ctaacagtcc agatgtacac ccagcctcag ggggaaggca gctctctcca
6421 gacagagtct cagggcccag caagggtcagg ttatctgctt tcattcaggg caacaaatga
6481 tacaaatggg gccagggagt ggcaaggcca tgggggtagg tgggggtgtc tttttctttt
6541 cataaagtaa caacagacga gactgagggt aaacatcaga aaaaaacctc tggaatgacc
6601 ttcctcattc caggaggccc tgggaataagg aagaggcttc tttctgaggg agctttgagg
6661 aatttttgaca gctgttgaca tgggatttgg gaaagggtgaa gctgtgactg gaggggcagg
6721 agatgggtcca agtgtccatc cagagatgag actcttagaa tcaaagtgtt cagcccagga
6781 agtcttgagg atcccacctt ctgtggccct gcaccttatg ggaagccatt aagggggctc
6841 atctaggaat tctggttaca gcccagtgtc catcccagcg tatgtgcct ctttagggca
6901 gcccgaagg ccagccagcc tgtactctgg gcaagagccc aaaatggcta ggaatgtttg
6961 actcccttaa tctcttcccc agctacagag gaatcttttc tctgcctggg ctcagaatgg
7021 gactgccaac tggctcattg gtgggagaca cagtatcctc aaacctgtgg ccactggcat
7081 gacagtgggtg ctctgtctcc ctgggtgaca cccaccctag gcttcctcct ggatgtgatg
7141 gggattgcca gagaggctct tagcataaaa ggcattaggt gggcattttt ctgtgtgccc
7201 ccaaaaagct ccatggaaac aggcacctgg tagctgcgga acaccgtgg acttgtgtat
7261 atggtcatag gctttgggaa gacaggacgt aaaggaaaaat gagagaaaca aaatgggtca
7321 gatagctttg gccacagccc caggcagcct ttggggccta tgacacttag tgcccttaga
7381 tgggatacat cttgcctcgg cccaagact cctccaactt acccgcccc tccagggcct
7441 gcacagctta gagaggctca cagcttgga aatgctaggg cttcatcaga ccactgactt
7501 gactcagtgt ttgttaaaat ggaaccactc cgttggcct actgtttctc tcctgtactt
7561 cttgtaatga tagttattta ttgactctgg tagcaggcag ttcctaaata aagatgggtt
7621 ctcaacctgt tggggaaaaa aaaaaaaaaa (SEQ ID NO:9)

```

Figure 8C

13/115

AX01 (Axonin-1 precursor, NM_005076)

MGTATRRKPHLLLVAAVALVSSSAWSSALGSQTTFGPVFEDQPL
SVLFPEESTEEQVLLACRARASPPATYRWKMNGTEMKLEPGSRHQLVGGNLVIMNPTK
AQDAGVYQCLASNPVGTVVSREAILRFGFLQEFKSKEERDPVKAHEGWGVMLPCNPPAH
YPGLSYRWLLNEFPNFIPTDGRHFVSQTTGNLYIARTNASDLGNYSCLATSHMDFSTK
SVFSKFAQLNLAAEDTRLFAPSIKARFPAETYALVGQQTLECFAGNPVPRIKWRKV
DGSLSPOWTTAEPTLQIPSVSFEDEGTYECEAENSKGRDTVQGRIIVQAQPEWLKVIS
DTEADIGSNLRWGCAAAGKPRPTVRWLRNGEPLASQNRVEVLAGDLRFSLSLSDSGM
YQCVAENKHGTIYASAEALAVQALAPDFRLNPVRRLI PAARGGEILIPCQPRAAPKAVV
LWSKGTEILVNSSRVTVTPDGTLLIRNISRSDEGKYTCFAENFMGKANSTGILSVRDA
TKITLAPSSADINLGDNLTLQCHASHDPTMDLTFTWTLDDFPIDFDKPGGHYRRTNVK
ETIGDLTILNAQLRHGGKYTCMAQTVVDSASKEATVLVRGPPGPPGGVVVRDIGDTTI
QLSWSRGFDNHSPIAKYTLQARTPPAGKWKQVRTNPANIEGNAETAQVLGLTPWMDYE
FRVIASNILGTGEPSPSSKIRTREAAPSVAPSGLSGGGGAPGELIVNWPMSREYQN
GDGFGYLLSFRRQGSTHWQTARVPGADAQYFVYSNESVRPYTPFEVKIRSYNRRGDGP
ESLTALVYSAEEDPRVAPTKVWAKGVSSSEMNVTWEPVQQDMNGILLGYEIRYWKAGD
KEAAADRVRTAGLDTSARVSGLHPNTKYHVTVRAYNRAGTGPASPSANATTMKPPRR
PPGNISWTFSSSSLSIKWDPVVPFRNESAVTGYKMLYQNDLHLTPTLHLTGKNWIEIP
VPEDIGHALVQIRTTGPGGDGIPAEVHIVRNGGTSMMVENMAVRPAPHPGTVISHSVA
MLILIGSLEL (SEQ ID NO:10)

Figure 8D

14/115

NROB2 (Nuclear hormone receptor, NM_021969)

```

1 gagctggaag tgagagcaga tccctaacca tgagcaccag ccaaccaggg gcctgcccac
61 gccagggagc tgcaagccgc cccgccattc tctacgcact tctgagctcc agcctcaagg
121 ctgtcccccg accccgtagc cgctgcctat gtaggcagca ccggcccgtc cagctatgtg
181 cacctcatcg cacctgccgg gaggccttgg atgttctggc caagacagtg gccttcctca
241 ggaacctgcc atccttctgg cagctgcctc ccaggacca gcggcggctg ctgcagggtt
301 gctggggccc cctcttctg cttgggttgg cccaagatgc tgtgaccttt gaggtggctg
361 agggcccggg gccagcata ctcaagaaga ttctgctgga ggagcccagc agcagtggag
421 gcagtggcca actgccagac agaccccagc cctccctggc tgcggtgcag tggcttcaat
481 gctgtctgga gtccttctgg agcctggagc ttagcccca ggaatatgcc tgctgaaag
541 ggaccatcct cttcaacccc gatgtgccag gcctccaagc cgcctccac attgggcacc
601 tgcagcagga ggctcactgg gtgctgtgtg aagtcctgga accctgggtg ccagcagccc
661 aaggccgcct gacccgtgtc ctctcacgg cctccaccct caagtccatt ccgaccagcc
721 tgcttgggga cctcttcttt cgcctatca ttggagatgt tgacatcgct ggccttcttg
781 gggacatgct tttgctcagg tgacctgttc cagcccaggc agagatcagg tgggcagagg
841 ctggcagtgc tgattcagcc tggccatccc cagaggtgac ccaatgctcc tggaggggca
901 agcctgtata gacagcactt ggctccttag gaacagctct tcaactagcc acaccccaca
961 ttggacttcc ttggtttgga cacagtgtc cagctgcctg ggaggctttt ggtggtcccc
1021 acagcctctg ggccaagact cctgtccctt cttgggatga gaatgaaagc ttaggtgct
1081 tattggacca gaagtcctat cgactttata cagaactgaa ttaagttatt gatttttgta
1141 ataaaaggta tgaacacta aaaaaaaaa (SEQ ID NO:11)

```

FIGURE 9A

NROB2 (Nuclear hormone receptor, NM_021969)

```

MSTSQPGACPCQGAASRPAILYALLSSSLKAVPRPRSRCLCRQH
RPVQLCAPHRTCREALDVLAKTVAFRLNLPSEFWQLPPQDQRRLLQGCWGPLFLLGLAQ
DAVTFEVAEAPVPSILKKILLEPSSSSGGSGQLPDRPQPSLAAVQWLQCCLESFWSLE
LSPKEYACLKGTILFNPDPGLQAASHIGHLQOEAHWVLCEVLEPWCFAAQGRLTRVL
LTASTLKSIPSTLLGDLFFRPIIGDVLDIAGLLGDMLLLR (SEQ ID NO:12)

```

FIGURE 9B

15/115

TM7SF1 (NM_003272)

```

1  cggcgcgatg  cgcggagacc  cccgcggggg  cggcggcgcc  cgtgagcccc  gatgaggccc
61  gagegtcccc  ggccgcgcgg  cagcgccccc  ggcccgatgg  agaccccgcc  gtgggaccca
121  gcccgcaacg  actcgctgcc  gcccacgctg  accccggccg  tgcccccta  cgtgaagctt
181  ggccctcacc  tcgtctacac  cgtgttctac  gcgctgctct  tcgtgttcac  ctacgtgcag
241  ctctggctgg  tgctgcgtta  ccgccacaag  cggctcagct  accagagcgt  cttcctcttt
301  ctctgcctct  tctgggcctc  cctgcggacc  gtccctctct  ctttctactt  caaagacttc
361  gtggcggcca  attcgctcag  ccccttcgtc  ttctggctgc  tctactgctt  cctgtgtgct
421  ctgcagtttt  tcacctcac  gctgatgaac  ttgtacttca  cgcaggatga  tttcaaagcc
481  aagtcaaaat  attctccaga  attactcaaa  taccggttgc  cctctacct  ggccctccct
541  ttcatcagcc  ttgttttctt  gttggtgaat  ttaacctgtg  ctgtgctggg  aaagacggga
601  aattgggaga  ggaaggttat  cgtctctgtg  cgagtggcca  ttaatgacac  gctcttcgtg
661  ctgtgtgccg  tctctctctc  catctgtctc  tacaaaatct  ctaagatgtc  cttagccaac
721  atttacttgg  agtccaagg  ctctccgtg  tgtcaagtga  ctgccatcgg  tgtcaccgtg
781  atactgcttt  acacctctcg  ggccctgctac  aacctgttca  tcctgtcatt  ttctcagaac
841  aagagcgtcc  attcctttga  ttatgactgg  tacaatgtat  cagaccaggc  agatttgaag
901  aatcagctgg  gagatgctgg  atacgtatta  tttggagtgg  tgttatttgt  ttgggaactc
961  ttacctacca  ccttagtcgt  ttatttcttc  cgagttagaa  atcctacaaa  ggaccttacc
1021  aaccttgaa  tggccccag  ccattggattc  agtcccagat  cttatttctt  tgacaaccct
1081  cgaagatatg  acagtgatga  tgaccttgcc  tgggaacattg  cccctcaggg  acttcaggga
1141  ggttttgctc  cagattacta  tgattgggga  caacaaacta  acagcttcct  ggcacaagca
1201  ggaactttgc  aagactcaac  tttggatcct  gacaaaccaa  gccttgggta  gcatcagtta
1261  acagttttat  ggacgattcc  tcagatgaaa  agcttcagaa  aagcatagt  acagctgaat
1321  ttttagggca  cttttcctta  agaaatagaa  cttgattttt  atttgttaca  ggtttccaat
1381  ggccccatag  gaataagcaa  taatgtagac  tgataaacc  ttattttagt  actaaaggag
1441  gagccttgct  atttcagtgg  gtataattta  aacttttta  agaaaactct  tacttttata
1501  aagatgtatt  ttgtataact  taaataataa  tgctaaagta  tactagggtt  ttttttctt
1561  gagaatgtta  ctgcaatcat  gttgtagttt  gcacagactt  ttatgcataa  ttcactttaa
1621  aaatatagaa  tatatggtct  aatagttttt  taaagctttt  ggactaaagt  attccacaaa
1681  tcttacctct  ttaggtcact  gatggtcact  ccgattctga  gtgccacatt  ggtagactcc
1741  taaaatacag  ttgacaactt  agccaattgc  aactccagt  ttgataatta  aaatgaaatg
1801  gtaaagcagc  agactgtaag  gtcttttagag  atttttttt  aaggttcagg  ccgtaggttc
1861  ctcaaggaa  ctcttaagtt  ttgccccaa  actggtactt  cttttcagta  gggcgcta  at
1921  gtatacacat  taatgataag  ttgataacat  taaaaatgta  gctgacttat  cctattaaac
1981  ctctctgct  atgttcac (SEQ ID NO:13)

```

FIGURE 10A

TM7SF1 (NM_003272)

```

MRPERPRPRGSAPGPMETPPWDPARNDLPPPTLTPAVPPYVKLG
LITVYTVFYALLFVFIYVQLWLVLRYRHKRLSYQSVFLFLCLFWASLRTVLFSFYFKD
FVAANSLSPFVFWLLYCFPVCLQFFTLTLMNLYFTQVIFKAKSKYSPELLKYRLPLYL
ASLFISLVFLLVNLTCAVLVKTGNWERKVIIVSVRVAINDTLFLVLCVSLISICLYKISK
MSLANIYLESKGSSVCQVTAIGVTVILLYTSRACYNLFILSFSQNKSVHSFDYDWINV
SDQADLKNQLGDAGYVLFVGVLFVWELLPTTLVVYFFRVRNPTKDLTNPGMVP SHGFS
PRSYFFDNPRRYDSDDDLAWNIPQGLQGGFAPDYDWDGQQTNSFLAQAGTLQDSTLD
PDKPSLG (SEQ ID NO:14)

```

FIGURE 10B

16/115

DLDH (dihydrolipamide dehydrogenase, NM_000108)

```

1  gcgcagggag gggagacctt ggcggacggc ggagccccag cggaggtgaa agtattggcg
61  gaaaggaaaa tacagcggaa aaatgcagag ctggagtcgt gtgtactgct ccttggccaa
121 gagaggccat ttcaatcgaa tatctcatgg cctacagggg ctttctgcag tgcctctgag
181 aacttacgca gatcagccga ttgatgctga tgtaacagtt ataggttctg gtccctggagg
241 atatgttgct gctattaaag ctgcccagtt aggcctcaag acagtctgca ttgagaaaaa
301 tgaaacactt ggtggaacat gcttgaatgt tgggtgtatt ccttctaagg ctttattgaa
361 caactctcat tattaccata tggcccattg aacagatttt gcatctagag gaattgaaat
421 gtccgaagtt cgcttgaatt tagacaagat gatggagcag aagagtactg cagtaaaagc
481 ttttaacaggt ggaattgccc acttattcaa acagaataag gttgttcatg tcaatggata
541 tggaaagata actggcaaaa atcaagtcac tgctacgaaa gctgatggcg gcactcaggt
601 tattgataca aagaacattc ttatagccac ggggttcagaa gttactcctt ttcctggaat
661 cacgatagat gaagatacaa tagtgctatc tacagggtgct ttatctttaa aaaaagttcc
721 agaaaagatg gttgttattg gtgcaggagt aataggtgta gaattggggt cagtttggca
781 aagacttggt gcagatgtga cagcagttga attttttaggt catgtagggt gagttggaat
841 tgatatggag atatctaaaa actttcaacg catccttcaa aaacaggggt ttaaatttaa
901 attgaatata aaggttactg gtgctaccaa gaagtcagat ggaaaaattg atgtttctat
961 tgaagctgct tctggtggtg aagctgaagt tatcacttgt gatgtactct tggtttgcct
1021 tggccgacga ccctttacta agaatttggg actagaagag ctgggaattg aactagatcc
1081 tagaggtaga attccagtca ataccagatt tcaaaactaaa attccaaata tctatgccat
1141 tgggtgatgta gttgctggtc caatgctggc tcacaaagca gaggatgaag gcattatctg
1201 tgttgaagga atggctgggt gtgctgtgca cattgactac aattgtgtgc catcagtgat
1261 ttacacacac cctgaagttg cttgggttgg caaatcagaa gagcagttga aagaagaggg
1321 tattgagtac aaagttggga aattcccatt tgctgctaac agcagagcta agacaaatgc
1381 tgacacagat ggcattggtg agatccttgg gcagaaatcg acagacagag tactggggagc
1441 acatatcttt ggaccaggtg ctggagaaat ggtaaatgaa gctgctcttg ctttgggaata
1501 tggagcatcc tgtgaagata tagctagagt ctgtcatgca catccgacct tatcagaagc
1561 ttttagagaa gcaaatcttg ctgctcatt tggcaaatca atcaactttt gaattagaag
1621 attatatatt tttttttctg aaatttctct ggagcttttg tagaagtcac attcctgaac
1681 aggatattct cacagctcca agaatttcta ggactgaatt atgaaacttt tgggaaggtat
1741 ttaatagggt tggacaaaat ggaatactct tatatctata ttttacataa atttagtatt
1801 ttgtttcagt gcactaatat gtaagacaaa aaggactact tattgtagtc atcctggaat
1861 atctccgtca actcatattt tcatgctgtt catgaaagat tcaatgcccc tgaattttaa
1921 tagctctttt ctctgataca gaaaagttga attttacatg gctggagcta gaatttgata
1981 tgtgaacagt tgtgtttgaa gcacagtgat caagttattt ttaatttggg tttcacattg
2041 gaaacaagtc agtcattcag aatgatttca aatgtctata aaccaaactg atgtaagtaa
2101 atgggtctct acttgtttta tttaacctct aaattctttc attttagggg tagcatttgt
2161 gttgaagagg ttttaaagct tccattgttg tctgcaactc tgaagggtaa ttatatagtt
2221 acccaaatta agagagtcta tttacggaac tcaaatacgt gggcattcaa atgtattaca
2281 gtgggggaatg aagatactga aataaacgtc ttaaatattc (SEQ ID NO:15)

```

FIGURE 11A

DLDH (dihydrolipamide dehydrogenase, NM_000108)

```

MQSWSRVYCSLAKRGHFNRISHGLQGLSAVPLRTYADQPIDADV
TVIGSGPGGYVAAIKAAQLGFKTVCI EKNETLGGTCLNVGCIPSKALLNNSHYHMAH
GTDFASRGIE MSEVRLNLDKMM EQKSTAVKALTGGIAHLFKQNKVVHVNGYGKITGKN
QVTATKADGGTQVIDTKN ILIATGSEVTPFPGITIDEDTIVSSTGALS LKKVPEKMOV
IGAGVIGVELGSVWQRLGADV TAVEFLGHVGGVGIDMEISKNFQRI LQKQGFKFLNT
KVTGATKKSDGKIDVSI EAASGGKAEVITCDVLLVCIGRRPFTKN LGLEELGIELDPR
GRIPVNTRFQTKIPNIY AIGDVVAGPMLAHKA EDEGIICVEGMAGGAVHIDYNCVPSV
IYTHPEVAWVGKSEEQLKEEG IEYKVGKFPFAANSRAKT NADTDGMVKILGQKSTDRV
LGAHILGPGAGEMVNEAAL ALEYGASCEDIARVCHAHPTL SEAFREANLAASF GKSIN
F (SEQ ID NO:16)

```

FIGURE 11B

17/115

MAT2B (methionine adenosyltransferase II, beta, NM_013283)

```

1 gttctgggcc taggggagggc gggccgaggg cgtctgagct gaggcccgcg tcgatacctgg
61 gttggaggag gtggcgggccg ctgaggctgc ggcgtgaaga cggcgggcat ggtggggcgg
121 gagaaagagc tctctataca ctttgttccc gggagctgtc ggctgggtga ggaggaagtt
181 aacatcccta ataggagggt tctggttact ggtgccactg ggcttcttgg cagagctgta
241 cacaaagaat ttcagcagaa taattggcat gcagttggct gtggtttcag aagagcaaga
301 ccaaaatttg aacaggttaa tctgttggat tctaatagcag ttcatacat cattcatgat
361 tttcagcccc atgttatagt acattgtgca gcagagagaa gaccagatgt tgtagaaaat
421 cagccagatg ctgcctctca acttaagtgt gatgcttctg ggaatttagc aaaggaagca
481 gctgctgttg gagcatttct catctacatt agctcagatt atgtatttga tggaaacaaat
541 ccaccttaca gagaggaaga cataccagct cccctaaatt tgtatggcaa aacaaaatta
601 gatggagaaa aggctgtcct ggagaacaat ctaggagctg ctgttttgag gattcctatt
661 ctgtatgggg aagttgaaaa gctcgaagaa agtgcgtgta ctgttatgtt tgataaagtg
721 cagttcagca acaagtcagc aaacatggat cactggcagc agaggttccc cacacatgtc
781 aaagatgtgg ccactgtgtg ccggcagcta gcagagaaga gaatgctgga tccatcaatt
841 aagggaaacct ttcactggctc tggcaatgaa cagatgacta agtatgaaat ggcactgtgca
901 attgcagatg ccttcaacct cccagcagc cacttaagac ctattactga cagccctgtc
961 ctagggcac aacgtccgag aaatgctcag cttgactgct ccaaattgga gaccttgggc
1021 attggccaac gaacaccatt tcgaattgga atcaaagaat cactttggcc tttcctcatt
1081 gacaagagat ggagacaaac ggtctttcat tagtttattt gtgttgggtt cttttttttt
1141 tttaaatgaa aagtatagta tgtggcactt tttaaagaac aaaggaaata gttttgtatg
1201 agtactttta ttgtgactct taggatcttt caggtaaagt atgctcttgc actagtgaat
1261 ttgtctaaag aaactaaagg gcagtcatgc cctgtttgca gtaatttttc tttttatcat
1321 tttgtttgtc ctggctaaac ttggagtttg agtatagtaa attatgatcc ttaaataattt
1381 gagagtcagg atgaagcaga tctgctgtag acttttcaga tgaaattggt cattctcgta
1441 acctccatat tttcaggatt tttgaagctg ttgacctttt catgttgatt attttaaatt
1501 gtgtgaaata gtataaaaaat cattgggtgt cattatttgc tttgcctgag ctcagatcaa
1561 aatgtttgaa gaaaggaact ttatttttgc aagttacgta cagtttttat gcttgagata
1621 tttcaacatg ttatgtatat tggaaacttct acagcttgat gcctcctgct tttatagcag
1681 tttatgggga gcacttgaaa gagcgtgtgt acatgtattt tttttctagg caaacattga
1741 atgcaaacgt gtattttttt aatataaata tataactgtc cttttcatcc catgttgccg
1801 ctaagtgata tttcatatgt gtgggttatac tcataataat gggccttgta agtcttttca
1861 ccattcatga ataataataa atatgtactg ctggcatgta atgcttagtt ttcttgattt
1921 tactttcttt tttaaatgta aggaccaaac ttctaaacta attgttcttt ttgtgcttta
1981 atttttaaaa attacattct tctgatgtaa catgtgatac atacaaaaga atatatgtta
2041 atatgtattg aaataaaaaca caataaaatt aaaaaaaaaa aaaaaaaaaa (SEQ ID
NO:17)

```

FIGURE 12A

MAT2B (methionine adenosyltransferase II, beta, NM_013283)

```

MVGREKELSIHFVPGSCRLVEEEVNIPNRRVLVTGATGLLGRAV
HKEFQQNNWHAVGCGFRRARPKFEQVNLLDSNAVHHI IHDFQPHVIVHCAAERRPDVV
ENQPDAAASQLNVDASGNLAKEAAVGAFLIYISSDYVFDGNTNPPYREEDI PAPLNLYG
KTKLDGEKAVLENNLGA AVLRIPILYGEVEKLEESAVTMFDKVQFSNKSANMDHWQQ
RFPTHVKDVATVCRQLAEKRMLDPSIKGTFHWSGNEQMTKYEMACAIADAFNLPSSHL
RPITDSPVLGAQRPRNAQLDCSKLETLGIGQRTPFRIKESLWPFLLIDKRWRQTVFH (SEQ ID
NO:18)

```

FIGURE 12B

18/115

STC-2 (stanniocalcin-2, NM_003714)

```

1 gaggaggagg gaaaaggcga gaaaaaagga agagtgggag gaggagggga agcggcgaaag
61 gaggaagagg aggaggagga agaggggagc acaaaggatc caggtctccc gacgggaggt
121 taataccaag aaccatgtgt gccgagcggc tgggccagtt catgaccctg gctttggtgt
181 tggccacctt tgacccggcg cgggggaccg acgccaccaa cccacccgag ggtccccaag
241 acaggagctc ccagcagaaa ggccgcctgt ccctgcagaa tacagcggag atccagcact
301 gtttgggtcaa cgctggcgat gtgggggtgtg gcgtgtttga atgtttcgag aacaactctt
361 gtgagattcg gggcttacat gggatttgca tgacttttct gcacaacgct ggaaaatttg
421 atgccagggg caagtcattc atcaaagacg ccttgaaatg taaggccac gctctgcggc
481 acaggttcgg ctgcataagc cggaaagtgc cggccatcag ggaaatggtg tcccagttgc
541 agcgggaatg ctacctcaag cacgacctgt gcgcggctgc ccaggagaac acccgggtga
601 tagtgagat gatccatttc aaggacttgc tgctgcacga accctacgtg gacctcgtga
661 acttgctgct gacctgtggg gaggaggtga aggaggccat caccacagc gtgcaggttc
721 agtgtgagca gaactgggga agcctgtgct ccatcttgag cttctgcacc tcggccatcc
781 agaagcctcc cacggcgccc ccgagcgcg agccccaggt ggacagaacc aagctctcca
841 gggcccacca cggggaagca ggacatcacc tcccagagcc cagcagtagg gagactggcc
901 gaggtgccaa gggtgagcga ggtagcaaga gccacccaaa cgcccatgcc cgaggcagag
961 tcgggggctt tggggctcag ggaccttcg gaagcagcga gtgggaagac gaacagtctg
1021 agtattctga tatccgagg tgaaatgaaa ggccctggcca cgaaatcttt cctccacgcc
1081 gtccattttc ttatctatgg acattccaaa acatttacca ttagagaggg gggatgtcac
1141 acgcaggatt ctgtggggac tgtggacttc atcgaggtgt gtgttcgcgg aacggacagg
1201 tgagatggag accctgggg ccgtggggtc tcaggggtgc ctggtgaatt ctgcacttac
1261 acgtactcaa gggagcgcgc ccgcgttatt ctcgtacctt tgtctctctt ccatctgtgg
1321 agtcagtggg tgcggccgc tctgttgtgg gggaggtgaa ccagggaggg gcagggcaag
1381 gcagggcccc cagagctggg ccacacagtg ggtgctgggc ctgcgccga agcttctggt
1441 gcagcagcct ctggtgctgt ctccgcggaa gtcagggcgg ctggattcca ggacaggagt
1501 gaatgtaaaa ataaatatcg cttagaatgc aggagaaggg tggagaggag gcaggggccc
1561 aggggggtgct tggtgccaaa ctgaaattca gtttcttggt tggggccttg cggttcagag
1621 ctcttggcga ggggtggagg aggagtgtca tttctatgtg taatttctga gccattgtac
1681 tgtctgggct gggggggaca ctgtccaagg gagtggcccc tatgagttta tattttaacc
1741 actgcttcaa atctcgattt cacttttttt atttatccag ttatatctac atatctgtca
1801 tctaaataaa tggctttcaa acaaagcaac tgggtcatta aaaccagctc aaagggggtt
1861 taaaaaaaaa aaaaccagcc catcctttga ggctgatttt tctttttttt aagttctatt
1921 taaaagcta tcaaacagcg acatagccat acatctgact gcctgacatg gactcctgcc
1981 cacttggggg aaaccttata ccagaggaa aatacacacc tggggagtag atttgacaaa
2041 tttcccttag gatttcgtta ctccaccttg accctcagcc aagattggta aagctcgtc
2101 ctggcgattc caggagacc agctggaaa cctggcttct catgtgaggg gatgggaaag
2161 gaaagaagag aatgaagact acttagtaat tcccatcagg aaatgctgac cttttacata
2221 aaatcaagga gactgctgaa aatctctaag ggacaggatt ttccagatcc taattggaaa
2281 tttagcaata aggagaggag tccaagggga caaataaagg cagagagaga gagagagaga
2341 gggagaggaa gaaaagagag agagaaaaga gcctcgtgcc (SEQ ID NO:19)

```

FIGURE 13A

STC-2 (stanniocalcin-2, NM_003714)

```

MCAERLQGFTLALVLATFDPARGTDATNPPEGPQDRSSQKGR
LSLQNTAEIQHCLVNAGDVGCVFECFENNSCEIRGLHGICMTFLHNAGKFDAQGKSF
IKDALCKKAHALRHRFGCISRKCPAIREMVSQLORECYLKHDLCAAAQENTRVIVEMI
HFKDLLLHEFPYVDLVNLLLTCEGEVKEAITHSVQVQCEQNWGSLCSILSFCTSAIQKP
PTAPPERQPQVDRTKLSRAHHGEAGHHLPEPSSRETGRGAKGERGSKSHPNHARGRV
GGLGAQGPSGSSEWEDEQSEYSDIRR (SEQ ID NO:20)

```

FIGURE 13B

19/115

PPBI (alkaline phosphatase, intestinal precursor, NM_001631)

```

1  gttcctggtg tccccacttc gcctccctcc tgctgcccc aagacatgca ggggccctgg
61  gtgctgctgc tgctgggcct gaggtacacg ctctccctgg gcgtcatccc agctgaggag
121  gagaacccgg ccttctggaa ccgccaggca gctgaggccc tggatgctgc caagaagctg
181  cagcccatcc agaaggtcgc caagaacctc atcctcttcc tgggcgatgg gttgggggtg
241  cccacggtga cagccaccag gatcctaaag gggcagaaga atggcaaact ggggcctgag
301  acgcccctgg ccatggaccg cttcccatac ctggctctgt ccaagacata caatgtggac
361  agacagggtg cagacagcgc agccacagcc acggcctacc tgtgcggggt caaggccaac
421  ttccagacca tcggcttgag tgcagccgcc cgctttaacc agtgcaacac gacacgcggc
481  aatgaggtca tctccgtgat gaaccgggcc aagcaagcag gaaagtcaat aggagtgggtg
541  accaccacac ggggtgcagca cgcctcgcca gccggcacct acgcacacac agtgaaccgc
601  aactgggtact cagatgctga catgcctgcc tcagcccgcc agggggggtg ccaggacatc
661  gccactcagc tcatctccaa catggacatt gacgtgatcc ttggcgaggg ccgcaagtac
721  atgtttccca tggggacccc agaccctgag taccagctg atgccagcca gaatggaatc
781  aggctggacg ggaagaacct ggtgcaggaa tggctggcaa agcaccaggg tgccctggtat
841  gtgtggaacc gactgagct catgcaggcg tccctggacc agtctgtgac ccatctcatg
901  ggcctctttg agcccggaga cacgaaatat gagatcctcc gagacccac actggacccc
961  tcctgatgg agatgacaga ggctgccctg cgctgctga gcaggaaccc ccgcggcttc
1021  tacctctttg tggagggcgg ccgcactcac catggtcatc atgaggggtg ggcttaccag
1081  gcagtcatg aggcggtcat gttcgacgac gccattgaga gggcgggcca gctcaccagc
1141  gaggaggaca cgctgacctt cgtcacccgt gaccactccc atgtcttctc ctttgggtggc
1201  tacaccttgc gagggagctc catcttcggg ttggccccc gcaaggctca ggacagcaaa
1261  gcctacacgt ccatcctgta cggcaatggc ccgggctacg tgttcaactc aggcgtgcga
1321  ccagacgtga atgagagcga gagcgggagc cccgattacc agcagcaggc ggcgggtgcc
1381  ctgtcgtccg agaccacagg aggcgaagac gtggcggtgt ttgcgcgcgg cccgcaggcg
1441  cacctggtgc atggtgtgca ggagcagagc ttcgtagcgc atgtcatggc cttcgtgcc
1501  tgtctggagc cctacacggc ctgogacctg gcgctcccc cctgcaccac cgacgccgcg
1561  caccagttg ccgcgtcgct gccactgctg gccgggacct tgctgctgct gggggcgctc
1621  gctgctccct gactgcccc ctcggagtt atcctgctcc ccactccgg gcgtcctgcc
1681  ctgttccccg tcctgagccg ccacttcag cgaacacaca caggtgtcct gccgttggac
1741  cttcacctcc tagagataaa ccagcctcag ctggcgagc ggggcccttc ttcctccgc
1801  atccccctca gggagcagga gccagggcg ccctgggagc tgagcctggg acttccagga
1861  cctccccctca ggttggtctc tgattcttcc tcccaacccc agagactgca gatttgtgcc
1921  atgcggctgc ctgcacccca gacaataaag ggaccaaacc caccacccc ccacctgcc
1981  tctatcctaa ggaagaccaa gcaggcctgg acccagagac gtcccccatc gtgggacacg
2041  acacaccag accgcgtgcc ccaccgtctt agcttcaatc ctggcagcac ctggtagacc
2101  caaggacttg ggtggatcag gacacctgaa gaagagaagc ttccggcaac ctgcaacc
2161  acccaaggag gctactggat cggggattcc caggggggct ttgacacagt cctctgctgt
2221  ctccccacta ggatcattcc acaccctgc acctgaccaa gggaccaatg aggcagaggc
2281  ttgccccaa gtcacagccac tcagatgctt cctgcccccc agtgcccatt ccaggtcacc
2341  agatccaagg agcgttgag gagctctggg tacagggcag caaccagag cccatgggcc
2401  ctcccgggac atctggatgc tgggcataga tttctcaaca aggaagactc cctgcctcc
2461  tcaaggtctc cattctccta ggagacaaag caataataaa aggtgttaga caatgt (SEQ

```

ID NO:21)

FIGURE 14A

PPBI (alkaline phosphatase, intestinal precursor, NM_001631)

```

MQGPWVLLLLLGLRLQLSLGVI PAEEENPAFWNRQAAEALDAKK
LQPIQKVAKNLILFLGDGLGVPTVTATRI LKGQKNGKLG PETPLAMDRFPYLALSKTY
NVDRQVPDSAATATAYLCGVKANFQTI GLSAAARFNQCNTTRGNEVISVMNRAKQAGK
SVGVVTTTRVQHASPAGTYAHTVNRNWYS DADMPASARQEGCQDIATQLISNMDIDVI
LG GGRKYMFPMPGTPDPEYPADASQNGIR LDGKNLVQEWLAKHQGAWYVWNRTELMQAS
LDQSVTHLMGLFEPGDTKYEILRDPTLDPS LMEMTEAALRLLSRNPRGFYLFVEGGRI
DHGHHEGVAYQAVTEAVMFDDAIERAGQLT SEEDTLTLVTADHSHVFSFGGYTLRGSS
IFGLAPSKAQDSKAYTSILYGNPGYVFNS GVRPDVNESESGSPDYQQQAAVPLSSET
HGGEDVAVFARGPQAHLVHGVQEQS FVAHVMAFAACLEPYTACDLALPACTTDAHPV
AASLPLLAGTLLLLGASAAP (SEQ ID NO:22)

```

FIGURE 14B

20/115

SLNAC1 (sodium channel receptor SLNAC1, NM_004769)

```

1  agaattcggc acgacggggt tctggccatg aagcccacct caggcccaga ggaggcccgg
61  cggccagcct cggacatccg cgtgttcgcc agcaactgct cgatgcacgg gctgggcccac
121 gtcttcgggc caggcagcct gagcctgcgc cgggggatgt gggcagcggc cgtgggtcctg
181 tcagtggcca ccttcctcta ccaggtggct gagaggggtg gctactacag ggagttccac
241 caccagactg ccctggatga gcgagaaagc caccggctca tcttcccggc tgtcaccctg
301 tgcaacatca acccactgcg ccgctcgcgc ctaacgcccc acgacctgca ctgggctggg
361 tctgcgctgc tgggcctgga tcccgcagag cagcgcgcct tctgcgcgc cctgggcccgg
421 cccctgcac cgcccggtt catgcccagt cccaccttg acatggcgca actctatgcc
481 cgtgctgggc actccctgga tgacatgctg ctggactgtc gcttcctgg ccaaccttgt
541 gggcctgaga acttcaccac gatcttcacc cggatgggaa agtgctacac atttaactct
601 ggcgctgatg gggcagagct gctcaccact actaggggtg gcatgggcaa tgggctggac
661 atcatgctgg acgtgcagca ggaggaatat ctacctgtgt ggagggacaa tgaggagacc
721 ccgtttgagg tggggatccg agtgcagatc cacagccagg aggagccgcc catcatcgat
781 cagctgggct tgggggtgtc cccgggctac cagaccttg tttcttgcca gcagcagcag
841 ctgagcttcc tgccaccgcc ctggggcgat tgcagttcag catctctgaa ccccaactat
901 gagccagagc cctctgatcc cctaggtccc cccagcccc gcccagccc tccctatacc
961 cttatggggg gtcgcctggc ctgcgaaacc cgctacgtgg ctcggaagtg cggctgcccga
1021 atgggtgtaca tgccaggcga cgtgccagtg tgcagcccc agcagtacaa gaactgtgcc
1081 cacccgcca tagatgccat gcttcgcaag gactcgtgcg cctgccccaa cccgtgcgcc
1141 agcacgcgct acgccaagga gctctccatg gtgcggatcc cgagccgcgc cgccgcgcgc
1201 ttcctggccc ggaagctcaa ccgcagcgag gcctacatcg cggagaacgt gctggccctg
1261 gacatcttct ttgaggccct caactatgag accgtggagc agaagaaggc ctatgagatg
1321 tcagagctgc ttggtgacat tggggggccag atggggctgt tcatcggggc cagcctgctc
1381 accatcctcg agatcctaga ctacctctgt gaggtgttcc gagacaaggt cctgggatat
1441 ttctggaacc gacagcactc ccaaaggcac tccagacca atctgttca ggaagggtg
1501 ggcagccatc gaaccgaagt tcccacctc agcctgggcc ccagacctc caccctccc
1561 tgtgccgtca ccaagactct ctccgcctcc caccgcacct gctacctgt cacacagctc
1621 tagacctgct gtctgtgtcc tgggagcccc gccctgacat cctggacatg cctagcctgc
1681 acgtagcttt tccgtcttca ccccaaataa agtcctaata catcaaaaaa aaaaaaaaaa
1741 aaaaaa (SEQ ID NO:23)

```

FIGURE 15A

SLNAC1 (sodium channel receptor SLNAC1, NM_004769)

```

MKPTSGPEEARRPASDIRVFASNCMSHGLGHVFGPGSLSLRRGM
WAAAVVLSVATFLYQVAERVYREFHHQTALDERESHRLIFPAVTLNINPLRRSRL
TPNDLHWAGSALLGLDPAEHAFLRALGRPPAPPGFMPSPTFDMAQLYARAGHSLDDM
LLDCRFRGQPCGPENFTTIFTRMGKCYTFNSGADGAELLTTTRGGMGNGLDIMLDVQQ
EEYLPVWRDNEETPFEVGIRVQIHSQEEPPIIDQLGLGVSPGYQTFVSCQQQQLSFLP
PPWGD CSSASLNP NYEPEPSDPLGSPSPSPSPPYTLMGCRLACETRYVARKCGCRMVY
MPGDV P V C S P Q Q Y K N C A H P A I D A M L R K D S C A C P N P C A S T R Y A K E L S M V R I P S R A A R F
L A R K L N R S E A Y I A E N V L A L D I F F E A L N Y E T V E Q K K A Y E M S E L L G D I G G Q M G L F I G A S L
L T I L E I L D Y L C E V F R D K V L G Y F W N R Q H S Q R H S S T N L L Q E G L G S H R T Q V P H L S L G P R P P
P P C A V T K T L A S H R T C Y L V T Q L (SEQ ID NO:24)

```

FIGURE 15B

21/115

CAH4 (carbonic anhydrase iv precursor, NM_000717)

```

1  ctcggtgctc gaccccggtc cagaggactc tttgctgtcc cgcaagatgc ggatgctgct
61  ggcgctcctg gccctctccg cggcgcgggc atcggccagt gcagagtcac actggtgcta
121 cgagggtcaa gccgagtcct ccaactaccc ctgcttggtg ccagtcaagt ggggtggaaa
181 ctgccagaag gaccgccagt ccccatcaa catcgtcacc accaaggcaa aggtggacaa
241 aaaactggga cgcttcttct tctctggcta cgataagaag caaacgtgga ctgtccaaaa
301 taacgggcac tcagtgatga tgttgctgga gaacaaggcc agcatttctg gaggaggact
361 gcctgccccca taccaggcca aacagttgca cctgcactgg tccgacttgc catataaggg
421 ctcgggagcac agcctcgatg gggagcactt tgccatggag atgcacatag tacatgagaa
481 agagaagggg acatcgagga atgtgaaaga ggcccaggac cctgaagacg aaattgcggt
541 gctggccttt ctggtggagg ctggaaccca ggtgaacgag ggcttccagc cactggtgga
601 ggcactgtct aatatcccca aacctgagat gagcactacg atggcagaga gcagcctgtt
661 ggacctgctc cccaaggagg agaaactgag gcactacttc cgctacctgg gctcactcac
721 cacaccgacc tgcgatgaga aggtcgtctg gactgtgttc cgggagccca ttcagcttca
781 cagagaacag atcctggcat tctctcagaa gctgtactac gacaaggaac agacagttag
841 catgaaggac aatgtcaggc ccctgcagca gctggggcag cgacggtga taaagtcgg
901 ggccccgggt cggccgctgc cctgggccct gcctgccctg ctgggccccca tgctggcctg
961 cctgctggcc ggcttctctg gatgatggct cacttctgca cgcagcctct ctggtgcctc
1021 agctctccaa gttccaggct tccggctcct agccttccca ggtgggactt taggcatgat
1081 taaaatatgg acatatTTTT ggag (SEQ ID NO:25)

```

FIGURE 16A

CAH4 (carbonic anhydrase iv precursor, NM_000717)

```

RMLLALLALSAAARPSASAESHWCYEVQAESSNYPCLVPVKWGG
CQKDRQSPINIVTTKAKVDKKLGRFFFSGYDKKQTTWTQNNGHSVMMLENKASISG
GLPAPYQAKQLHLHWSLDLPYKGSEHSLDGEHFAMEMHIVHEKEKGTSRNVKEAQDPE
EIAVLAFIVEAGTQVNEGFQPLVEALSNI PKPEMSTTMAESSLLDLLPKKEKLRHYF
YLGLSLTTPTCDEKVVTVFREPIQLHREQILAFSQKLYYDKEQTVSMKDNVRPLQQL
QRTVIKSGAPGRPLPWALPALLGPMLACLLAGFLR (SEQ ID NO:26)

```

FIGURE 16B

22/115

PA21 (phospholipase a2 precursor, NM_000928)

```
1  tggtcacatctc agttcttttc tcaccttgac tgcaagatga aactccttgt gctagctgtg
61  ctgctcacag  tggccgccgc cgacagcggc atcagccctc gggccgtgtg gcagttccgc
121 aaaatgatca agtgcgatgat cccggggagt gaccccttct tggaatacaa caactacggc
181 tgctactgtg gcttgggggg ctcaggcacc cccgtggatg aactggacaa gtgctgccag
241 acacatgaca actgctatga ccaggccaag aagctggaca gctgtaaatt tctgctggac
301 aacccgtaaca cccacaccta ttcatactcg tgctctggct cggcaatcac ctgtagcagc
361 aaaaacaaaag agtgtgaggc cttcatttgc aactgcgacc gcaacgctgc catctgcttt
421 tcaaaagctc catataacaa ggcacacaag aacctggaca ccaagaagta ttgtcagagt
481 tgaatatcac ctctcaaaag catcacctct atctgcctca tctcacactg tactctccaa
541 taaagcacct tgttgaaaga cctcaaaaaa aaaaaaaaaa aaaaa (SEQ ID NO:27)
```

FIGURE 17A

PA21 (phospholipase a2 precursor, NM_000928)

```
KLLVLAVLLTVAAADSGISPRAVWQFRKMIKCVIPGSDPFLEY
NYGCYCGLGGSGTPVDELDKCCQTHDNCYDQAKKLDCKFLLDNPYHTYSSCSGS
ITCSSKNKECEAFICNCDRNAAICFSKAPYNKAHKNLDTKKYCQS (SEQ ID NO:28)
```

FIGURE 17B

23/115

PAR2 (proteinase activated receptor 2 precursor, NM_005242)

```

1  tgaaacctaa cccgccctgg ggaggcgcg cagcagaggct ccgattcggg gcaggtgaga
61  ggctgacttt ctctcggtgc gtccagtgga gctctgagtt tcgaatcggc ggcggcggat
121  tccccgcgcg cccggcgctc gggcttccag gaggatgcgg agccccagcg cggcgtggct
181  gctggggggc gccatcctgc tagcagcctc tctctcctgc agtggcacca tccaaggaac
241  caatagatcc tctaaaggaa gaagccttat tggttaagggt gatggcacat cccacgtcac
301  tggaaaagga gttacagttg aaacagtcct ttctgtggat gagttttctg catctgtcct
361  cactggaaaa ctgaccactg tcttccttcc aattgtctac acaattgtgt ttgtgggtgg
421  tttgccaaagt aacggcatgg ccctgtgggt ctttcttttc cgaactaaga agaagcacc
481  tgctgtgatt tacatggcca atctggcctt ggctgacctc ctctctgtca tctgggtccc
541  cttgaagatt gcctatcaca tacatggcaa caactggatt tatggggaag ctctttgtaa
601  tgtgcttatt ggctttttct atggcaacat gtactgttcc attctcttca tgacctgcct
661  cagtgtgcag aggtattggg tcatcgtgaa ccccatgggg cactccagga agaaggcaaa
721  cattgccatt ggcactctcc tggcaatatg gctgctgatt ctgctggtca ccatcccttt
781  gtatgtcgtg aagcagacca tcttcattcc tgcctgaac atcacgacct gtcattgatg
841  tttgcctgag cagctcttgg tgggagacat gttcaattac ttctctctc ttgccattgg
901  ggtctttctg ttcccagcct tctcacagc cctgcctat gtgctgatga tcagaatgct
961  gcatcttct gccatggatg aaaactcaga gaagaaaagg aagaggggcca tcaaactcat
1021  tgtcactgtc ctggccatgt acctgatctg cttcactcct agtaaccttc tgccttggtg
1081  gcattatttt ctgattaaga gccaggggcca gagccatgtc tatgccctgt acattgtagc
1141  cctctgcctc tctaccctta acagctgcat cgaccctttt gtctattact ttgtttcaca
1201  tgatttcagg gatcatgcaa agaacgctct ctttgccga agtgtccgca ctgtaaagca
1261  gatgcaagta tccctcacct caaagaaaca ctccaggaaa tccagctctt actcttcaag
1321  ttcaaccact gtttaagacct cctattgagt ttccagggtc ctcatagggg aattgcacag
1381  taggatgtgg aacctgttta atgttatgag gacgtgtctg ttatttctta atcaaaaagg
1441  tctcaccaca taccatgtgg atgcagcacc tctcaggatt gctaggagct cccctgtttg
1501  catgagaaaa gtagtcccc aaattaacat cagtgtctgt ttcagaatct ctctactcag
1561  atgaccccg aaactgaacc aacagaagca gacttttcag aagatgggtg agacagaaac
1621  ccagtaactt gcaaaaagta gacttggtgt gaagactcac ttctcagctg aaattatata
1681  tatacacata tatatatatt acatctggga tcatgataga cttgttaggg cttcaaggcc
1741  ctgagagatg atcagtccea ctgaacgacc ttacaaatga ggaaaccaag ataaatgagc
1801  tgccagaatc aggttttcaa tcaacagcag tgagttggga ttggacagta gaatttcaat
1861  gtccagttag tgaggttctt gtaccacttc atcaaaatca tggatcttgg ctgggtgcgg
1921  tgctcatgct ctgtaatcct agcactttgg gaggtgagg caggcaatca cttgaggtca
1981  ggagttcgag accagcctgg ccatcatggc gaaacctcat ctctactaaa aatacaaaa
2041  ttaaccaggt gtgtggtgca cgtttgtaat ccagttact caggaggctg aggcacaaga
2101  attgagtatc actttaactc aggaggcaga ggttgacagt agccgagatt gcaccactgc
2161  actccagctt gggtgataaa ataaaataaa atagtctgta atcttggtca aaatgcagat
2221  tcctcagatt caataatgag agctcagact gggaacaggg cccaggaatc tgtgtggtac
2281  aaacctgcat ggtgtttatg cacacagaga tttgagaacc attgtttctg atgctgcttc
2341  catttgacaa agtgccgtga taatttttga aaagagaagc aaacaatggt gtctctttta
2401  tgctcagctt ataataaat ctgtttgttg acttattagg actttgaatt attctcttat
2461  taacctctg agtttttcta tgtattatta ttaaagaaaa atgcaatcag gattttaaac
2521  atgtaaatac aaatttttga taacttttga tgacttcagt gaaattttca ggtagtctga
2581  gtaatagatt gttttgccac ttagaatagc atttgccact tagtatttta aaaaataatt
2641  gttggagtat ttattgtcag ttttggtcac ttgttatcta atacaaaatt ataaagcctt
2701  cagagggttt ggaccacatc tcttggaata atagtttgca acatatttaa gagatacttg
2761  atgcaaaaat gactttatac aacgattgta tttgtgactt taaaaataa ttattttatt
2821  gtgtaattga tttataaata acaaaatttt ttttacaact taaaaaaaa aaaaaa (SEQ

```

ID NO:29)

FIGURE 18A

24/115

PAR2 (proteinase activated receptor 2 precursor, NM_005242)

RSPSAAWLLGAAILLAASLSCSGTIQGTNRSSKGRSLIGKVDG
SHVTGKGVTVETVFSVDEFSASVLTGKLTTVFLPIVYTIVFVVG LPSNGMALWVFLF
TKKKHPAVIYMANLALADLLSVIWFPLKIAYHIHGNNWIYGEALCNVLIGFFYGNMY
SILFMTCLSVQRYWVIVNPMGHSRKKANIAIGISLAIWLLILLVTIPLYVVKQTIFI
ALNITTCHDVLPEQLLVGDMFNYFLSLAIGVFLFPAFLTASAYVLMIRMLRSSAMDE
SEKKRKRAIKLIVTVLAMYLICFTPSNLLL VVHYFLIKSQGQSHVYALYIVALCLST
NSCIDPFVYYFVSHDFRDHAKNALLCRSVRTVKQM QVSLTSKKHSRKSSSYSSSSTT
KTSY (SEQ ID NO:30)

FIGURE 18B

25/115

IDE (insulin-degrading enzyme, NM_004969)

```

1  ccggctcgaa gcgcaacgag gaagcggtttg cggatgatccc ggcgactgcg ctggctaatag
61  cggtagccggc tagcgtgggt tctgcacccc gcactgcccgc gcaccttccg ctcagtcctc
121  ggccgcccggc tgccgcctcc ggagcgccctg tgtgggtttcc aaaaaaagac ttacagcaaa
181  atgaataatc cagccatcaa gagaatagga aatcacatta ccaagtctcc tgaagacaag
241  cgagaatatc gagggctaga gctggccaat ggtatcaaag tacttcttat gagtgatccc
301  accacggata agtcatcagc agcacttgat gtgcacatag gttcattgtc ggatcctcca
361  aatattgctg gcttaagtca tttttgtgaa catatgcttt ttttgggaac aaagaaatac
421  cctaaagaaa atgaatacag ccagtttctc agtgagcatg caggaagttc aaatgccttt
481  actagtggag agcataccaa ttactatttt gatgtttctc atgaacacct agaaggtgcc
541  ctagacaggt ttgcacagtt ttttctgtgc cccttggtcg atgaaagttg caaagacaga
601  gaggtgaatg cagttgattc agaacatgag aagaatgtga tgaatgatgc ctggagactc
661  tttcaattgg aaaaagctac agggaaatcct aaacaccctc tcagtaaatt tgggacaggt
721  aacaaatata ctctggagac tagaccaaac caagaaggca ttgatgtaag acaagagcta
781  ctgaaattcc attctgctta ctattcatcc aacttaatgg ctgtttgtgt tttaggtcga
841  gaatcttttag atgacttgac taatctggtg gtaaagttat tttctgaagt agagaacaaa
901  aatgttccat tgccagaatt tcctgaacac cttttccaag aagaacatct taacaactt
961  tacaaaatag taccatttaa agatattagg aatctctatg tgacatttcc catacctgac
1021  cttcagaaat actacaaatc aaatcctggt cattatcttg gtcattctcat tgggcatgaa
1081  ggtcctggaa gtctgtttatc agaacttaag tcaaagggct gggttaatac tcttggtggt
1141  gggcagaagg aaggagcccg aggttttatg ttttttatca ttaatgtgga cttgaccgag
1201  gaaggattat tacatgttga agatataatt ttgcacatgt ttcaatacat tcagaagtta
1261  cgtgcagaag gacctcaaga atgggttttc caagagtgca aggacttgaa tgctgttgct
1321  tttaggttta aagacaaaaga gagggcacgg ggctatacat ctaagattgc aggaatattg
1381  cattattatc ccctagaaga ggtgctcaca gcggaatatt tactggaaga atttagacct
1441  gacttaatag agatggttct cgataaactc agaccagaaa atgtccgggt tgccatagtt
1501  tctaaatctt ttgaaggaaa aactgatcgc acagaagagt ggtatggaaac ccagtacaaa
1561  caagaagcta taccggatga agtcatcaag aaatggcaaa atgctgacct gaatgggaaa
1621  tttaaacttc ctacaaagaa tgaatttatt cctacgaatt ttgagatttt accgttagaa
1681  aaagaggcga caccataccc tgctcttatt aaggatacag tcatgagcaa actttgggtc
1741  aaacaagatg ataagaaaaa aaagccgaag gcttgtctca actttgaatt tttcagccca
1801  tttgcttatg tggaccctt gcactgtaac atggcctatt tgtaccttga gctcctcaaa
1861  gactcactca acgagtatgc atatgcagca gagctagcag gcttgagcta tgatctccaa
1921  aataccatct atgggatgta tctttcagtg aaaggttaca atgacaagca gccaatatta
1981  ctaaagaaga ttattgagaa aatggctacc tttgagattg atgaaaaaag atttgaaatt
2041  atcaaagaag catatatgcg atctcttaac aatttccggg ctgaacagcc tcaccagcat
2101  gccatgtact acctccgctt gctgatgact gaagtggcct ggactaaaga tgagttaaaa
2161  gaagctctgg atgatgtaac ccttcctcgc cttaaggcct tcatacctca gctcctgtca
2221  cggctgcaca ttgaagccct tctccatgga aacataacaa agcaggctgc attaggaatt
2281  atgcagatgg ttgaagacac cctcattgaa catgctcata ccaaacctct ccttccaagt
2341  cagctggttc ggtatagaga agttcagctc cctgacagag gatgggtttg ttatcagcag
2401  agaaatgaag ttcacaataa ctgtggcatc gagatatact accaaacaga catgcaaagc
2461  acctcagaga atatgtttct ggagctcttc tgtcagatta tctcggaacc ttgcttcaac
2521  accctgcgca ccaaggagca gttgggctat atcgtcttca gcgggccacg tcgagctaatt
2581  ggcatacaga gcttgagatt catcatccag tcagaaaagc cacctcacta cctagaaaagc
2641  agagtggaa gctttcttaac taccatggaa aagtccatag aggacatgac agaagaggcc
2701  ttccaaaaac acattcagggc attagcaatt cgtcgactag acaaaccaaa gaagctatct
2761  gctgagtgtg ctaaatactg gggagaaaatc atctcccagc aatataattt tgacagagat
2821  aacactgagg ttgcatattt aaagacactt accaagggaag atatcatcaa attctacaag
2881  gaaatgttgg cagtagatgc tccaaggaga cataaggat ccgtccatgt tcttgccagg
2941  gaaatggatt cttgtcctgt tgttgagag ttcccatgtc aaaatgacat aaatttgtca
3001  caagcaccag ccttgccaca acctgaagtg attcagaaca tgaccgaatt caagcgtggt
3061  ctgccactgt ttccccctgt gaaaccacat attaacttca tggctgcaaa actctgaaga
3121  ttccccatgc atgggaaagt gcaagtggaat gcatctctga gtcttcaga gcctaagaaa
3181  atcatcttgg ccactttaat agtttctgat tcactattag agaaacaaac aaaaaattgt
3241  caaatgtcat tatgtagaaa tattataaat ccaaagtaa (SEQ ID NO:31)

```

FIGURE 19A

26/115

IDE (insulin-degrading enzyme, NM_004969)

MRYRLAWLLHPALPSTFRSVLGARLPPPERLCGFQKKTYSKMNN
PAIKRIGNHITKSPEDKREYRGLELANGIKVLLMSDPTTDKSSAALDVHIGSLSDPPN
IAGLSHFCEHMLFLGTTKYPKENEYSQFLSEHAGSSNAFTSGEHTNYYFDVSHEHLEG
ALDRFAQFFLCPLFDESCKDREVNADVSEHEKNVMNDAWRLFQLEKATGNPKHPFSKF
GTGNKYTLETRPNQEGIDVRQELLKFHSAYYSSNLMAVCVLGRESLDDLTNLVVKLFS
EVENKNVPLPEFPEHPFQEEHLKQLYKIVPIKDIRNLYVTFFPIPDLQKYYKSNPGHYL
GHLIGHEGPGSLLSELKSKGWNTLVGGQKEGARGFMFFIINVDLTEEGLLHVEDIIL
HMFQYIQKLRAEGPQEWVFQECKDLNAVAFRFKDKERPRGYTSKIAGILHYYPLEEV
TAEYLLLEEFRPDLIEMVLDKLRPENVRVAIVSKSFEGKTDRTTEWYGTQYKQEAIPDE
VIKKWQNADLNGKFKLPTKNEFIPTNFEILPLEKEATPYPALIKDTVMMSKLWFKQDDK
KKKPKACLNFEFFSPFAYVDPLHCNMAYLYLELLKDSLNEYAYAAELAGLSYDLQNTI
YGMYSVKGYNDKQPILLKKIIEKMATFEIDEKRFEIIEAYMRSLNNFRAEQPHQHA
MYYLRLLMTEVAWTKDELKEALDDVTLPRLKAFIPQLLSRLHIEALLHGNTTKQAALG
IMQMVEDTLIEHAHTKPLPSQLVRYREVQLPDRGWVFVYQQRNEVHNNCGIEIYYQTD
MQSTSENMFLELFCQIISEPCFNTLRTKEQLGYIVFSGPRRANGIQSLRFIIQSEKPP
HYLESRVEAFLITMEKSIEDMTEEAFQKHIALAIRRLDKPKKLSAECACYWGEIISQ
QYNFDRDNTEVAYLKTTLTKEDIIFKYKEMLAVDAPRRHKVSVHVLAREMDSCPVVGEF
PCQNDINLSQAPALPQPEVIQNMTEFKRGLPLFPLVKPHINFMAAKL (SEQ ID NO:32)

FIGURE 19B

27/115

MYO1A (myosin-1A, NM_005379)

```

1  cagggagcct  gggctggaag  aggcagcaaa  agggaaaatc  agaagagtgg  acactggcaa
61  gaggaggggca  gcctttttcc  cagcttcctt  gcaccatgga  cagctcccat  taagccacct
121  ctccatcctg  gggccaggac  tcttatgccc  cattcctgtc  aaattgagat  ttcattccacc
181  attctccaag  gacagtgaag  ttatacccta  gttccagtgt  tgggatcagt  ggccccctctg
241  gacatgcctc  tcctggaagg  ttctgtgggg  gtggaggatc  ttgtcctcct  ggaacccttg
301  gtggaggagt  cactgctcaa  gaatcttcag  cttcgctatg  aaaacaagga  gatttatacc
361  tacattggga  atgtggtgat  ctcagtgaat  ccctatcaac  agcttcccat  ctatgggcca
421  gagttcattg  ccaaataatca  agactatact  ttctatgagc  tgaagcccca  tatctacgca
481  ttggcaaatg  tggcgtaacca  gtcactgagg  gacagggacc  gagaccagtg  tatcctcatc
541  acaggcgaga  gtggatcagg  gaagactgag  gccagcaagc  tggatgatgtc  ttatgtggct
601  gccgtctgtg  ggaaaggaga  gcaggtgaac  tctgtgaagg  agcagctgct  acagtctaac
661  ccagtgtctg  aggccttttg  caatgccaa  accattcgca  acaacaattc  ctcccgattt
721  ggaaaataca  tggatattga  atttgacttc  aagggatccc  ccctcggtag  tgtcatcaca
781  aactatctgc  ttgagaaatc  ccgattagt  aagcagctca  aaggagaaag  gaacttccac
841  atcttctatc  agctgctggc  tggagcagat  gaacagctgc  tgaaggccct  gaagcttgag
901  cgggatacaa  ctggctatgc  ctatctgaat  catgaagtat  ccagagtgga  tggcatggac
961  gacgcctcca  gcttcagggc  tgtacagagt  gcaatggcag  tgattgggtt  ctccgaggag
1021  gagattcgac  aagtgtctaga  ggtgacatcc  atggtgctaa  agctggggaa  cgtgtgggtg
1081  gctgatgagt  tccaggccag  tgggatacca  gcaagtggca  tccgtgatgg  gagaggtgtt
1141  cgggagattg  gggagatggt  gggcttgaat  tcagaagaag  tagagagagc  tttgtgctcg
1201  aggaccatgg  aaacagccaa  ggaaaagggt  gtcactgcac  tgaatgttat  gcaggctcag
1261  tatgctcggg  acgccctggc  taagaacatc  tacagccgcc  tctttgactg  gatagtgaat
1321  cgaatcaatg  agagcatcaa  ggtgggcac  ggggaaaaga  agaaggtaat  gggagtcctt
1381  gatattctac  gttttgagat  attagaggat  aatagctttg  agcaatttgt  gatcaactac
1441  tgcaatgaga  agctgcagca  ggtgttcata  gagatgacct  tgaaagaaga  gcaagaggaa
1501  tataagagag  aaggcatacc  gtggacaaag  gtggactact  ttgataatgg  catcatttgt
1561  aagctcattg  agcataatca  gcgaggtatc  ctggccatgt  tggatgagga  gtgcctgcgg
1621  cctgggggtg  tcagtgaact  cactttccta  gcaaagctga  accagctctt  ctccaagcat
1681  ggccactacg  agagcaaagt  caccagaat  gccagcgtc  agtatgacca  caccatgggc
1741  ctgagctgct  tccgcatctg  ccactatgcg  ggcaagggtg  catacaacgt  gaccagcttt
1801  attgacaaga  ataatgacct  actcttccga  gacctgttgc  aggccatgtg  gaaggcccag
1861  caccctctcc  ttcggtcctt  gtttctctg  ggcaatccta  agcaggcatc  tctcaaacgc
1921  ccccgactg  ctggggccca  gttcaagagt  tctgtggcca  tcctcatgaa  gaattctgat
1981  tccaagagcc  ccaactacat  caggtgcata  aagcccaatg  agcatcagca  gcgaggtcag
2041  ttctcttcag  acctggtggc  aacctaggct  cggtaacctg  gactgctgga  gaacgtacgg
2101  gtgcgacggg  caggctatgc  ccaccgccag  ggttatgggc  ccttctctgga  aaggtaaccga
2161  ttgctgagcc  ggagcacctg  gcctcactgg  aatgggggag  accgggaagg  tgttgagaag
2221  gtcctggggg  agctgagcat  gtctcggggg  gagctggcct  ttggcaagac  aaagatcttc
2281  attagaagcc  ccaagactct  tttctacctc  gaagaacaga  ggcgcctgag  actccagcag
2341  ctggccacac  tcatacagaa  gatttaccga  ggctggcgct  gccgcaccca  ctaccaactg
2401  atgcgaaaga  gtcagatcct  catctcctct  tggtttcggg  gaaacatgca  aaagaaatgc
2461  tatgggaaga  taaaggcatc  cgtgttattg  atccaggctt  ttgtgagagg  gtggaaggcc
2521  cgaaagaatt  atcgcaaata  tttccgggtc  gaggctgccc  tcaccttggc  agatttcac
2581  tacaagagca  tggtagagaa  attcctactg  gggctgaaga  acaatttggc  atccacaaac
2641  gtcttagaca  agacatggcc  agccgcccc  tacaagtgcc  tcagcacagc  aaatcaggag
2701  ctgcagcagc  tcttctacca  gtggaagtgc  aagaggttcc  gggatcagct  gtccccgaag
2761  caggtagaga  tcctgagggg  aaagctctgt  gccagtgaac  tgttcaaggg  caagaaggct
2821  tcatatcccc  agagtgtccc  cattccattc  tgtggtgact  acattgggct  gcaagggaac
2881  cccaagctgc  agaagctgaa  aggcggggg  gaggggcctg  ttctgatggc  agaggccgtg
2941  aagaaggctc  atcgtggcaa  tggcaagact  tcttctcgga  ttctcctcct  gaccaagggc
3001  catgtgattc  tcacagacac  caagaagtcc  caggccaaaa  ttgtcattgg  gctagacaat
3061  gtggctgggg  tgtcagtcac  cagcctcaag  gatgggctct  ttagcttgca  tctgagttag
3121  atgtcatcgg  tgggctccaa  gggggacttc  ctgctggtca  gcgagcatgt  gattgaactg
3181  ctgacaaaa  tgtaccgggc  tgtgctggat  gccacgcaga  ggcagcttac  agtcaccgtg

```

FIGURE 20A

28/115

```
3241 actgagaagt tctcagtgag gttcaaggag aacagtgtgg ctgtcaaggt cgtccagggc
3301 cctgcagggtg gtgacaacag caagctacgc tacaaaaaaaa aggggagtca ttgcttggag
3361 gtgactgtgc agtgaggagg gggcaccatg cagagatggc agttgcttcc tcctgaacca
3421 gcactaatcc ccctctgccc tcctgtgtgg gaggatctct aaccctctg atcgtggcgc
3481 atggcttggg gattaaacta cccttgaaga ggacccttgt cccaaaccct tcttgttctc
3541 tcctccaaaa gtagcttcct ccaaccgcga gcctctctgc acactaataa aacatgtggc
3601 ttggaaaggt tcaaaaaaaaa aaaa (SEQ ID NO:33)
```

FIGURE 20B

29/115

MYO1A (myosin-1A, NM_005379)

PLLEGSGVEDLVLLLEPLVEESLLKKNLQRLRYENKEIYTYIGNV
ISVNPYQQLPIYGPEFIAKYQDYTFYELKPHIYALANVAYQSLRDRDRDQCILITGE
GSGKTEASKLVMSYVAAVCGKGQVNSVKEQLLQSNPVLEAFGNAKTIRNNNSSRFG
YMDIEFDFKGSPLGGVITNYLLEKSRLVKQLKGERNFHIFYQLLAGADEQLLKALKL
RDTTGYAYLNHEVSRVDGMDDASSFRAVQSAMAVIGFSEEEIRQVLEVTSMLVCLKGN
LVADEFQASGIPASGIRDGRGVREIGEMVGLNSEEVERALCSRTMETAKEKVVTALN
MQAQYARDALAKNIYSRLFDWIVNRINESIKVGIGEKKKVMGVLDIYGFEILEDNSF
QFVINYCNEKLQQVFIEMTLKEEQEEYKREGIPWTKVDYFDNGIICKLIEHNQRGIL
MLDEECLRPGVVS DSTFLAKLNQLFSKHGHYESKVTQNAQRQYDHTMGLSCFRICHY
GKVTYNVTSFIDKNNDLLFRDLLQAMWKAQHPLLRSLFPEGNPKQASLKRPPTAGAQ
KSSVAILMKNLYSKSPNYIRCIKPNEHQQRGQFSSDLVATQARYLGLENVRVRRAG
AHRQGYGPFLERYRLLSRSTWPHWNGGDREGVEKVLGELSMSSGELAFGKTKIFIRS
KTLFYLEEQRRLRLQQLATLIQKIYRGWRCRTHYQLMRKSQILISSWFRGNMQKKCY
KIKASVLLIQAFVRGWKARKNYRKYFRSEAALTADFIYKSMVQKFLGLKNNLPST
VLDKTWPAAPYKCLSTANQELQQLFYQWKCKRFRDQLSPKQVEILREKLCASELFKG
KASYPQSVPIPFCDYIGLQGNPKLQKLKGGEEGPVLMAEAVKKVNRGNGKTSSRIIL
LTKGHVILTDTKKSQAKIVIGLDNVAGSVTSLKDGLFSLHLSEMSSVGSKGDFLLV
EHVIELLTKMYRAVL DATQRQLTVTVTEKFSVRFKENSVAVKVVGPGAGDNSKLRY
KKGSHCLEVTVQ (SEQ ID NO:34)

FIGURE 20C

30/115

CYP2J2 (cytochrome P450 monooxygenase, NM_000775)

```

1 gagccatgct cgcggcgatg ggctctctgg cggetgccct ctgggcagtg gtccatcctc
61 ggactctcct actgggcact gtcgcctttc tgctcgctgc tgactttctc aaaagacggc
121 gccc aaagaa ctaccgcgcg gggccctggc gcctgccctt ccttggcaac ttcttccttg
181 tggacttcga gcagtcgcac ctggaggttc agctgtttgt gaagaaatat gggaaccttt
241 ttagcttgga gcttggtgac atatctgcag ttcttattac tggcttgccc ttaatcaaag
301 aagcccttat ccacatggac caaaactttg ggaaccgccc cgtgaccctt atgcgagaac
361 atatctttta gaaaaatgga ttgattatgt caagtggcca ggcatggaag gagcaaagaa
421 ggttcactct gacagcacta aggaactttg gtttaggaaa gaagagctta gaggaacgca
481 ttcaggagga ggcccaacac ctactgaag caataaaaga ggagaacgga cagccttttg
541 accctcattt caagatcaac aatgcagttt ccaatatcat ttgctccatc accttcggag
601 aacgctttga gtaccaggat agttggtttc agcagctgct gaagttacta gatgaagtca
661 catacttgga ggcttcaaag acatgccagc tctacaatgt ctttccatgg ataatgaaat
721 tcttgcttgg accccaccaa actctcttca gcaactggaa aaaactgaaa ttgtttgttt
781 ctcatatgat tgacaaaacac agaaaggatt ggaatcctgc agaaacaaga gactttattg
841 atgcttacct taaagaaatg tcaaagcaca caggcaatcc tacttcaagt ttccatgaag
901 aaaacctcat ctgcagcacc ctggacctct tctttgccgg aaccgagaca acttccacaa
961 ctctgcgatg ggctctgctt tatatggccc tctaccaga aatccaagaa aaagtacaag
1021 ctgagattga cagagtgatt ggccaggggc agcagccgag cacagccgcc cgggagttcca
1081 tgccctacac caatgctgtc atccatgagg tgcagagaaat gggcaacatc atccccctga
1141 acgttcccag ggaagtgaca gttgatacca ctttggctgg gtaccacctg cccaagggtg
1201 ccatgatcct gaccaatttg acggcgctgc acagggaccc cacagagtgg gccacccctg
1261 acacattcaa tccggaccat tttctggaga atggacagtt taagaaaagg gaagccttta
1321 tgccctttctc aataggaaaag cgggcacgct tcggagaaca gttggccagg actgagctgt
1381 ttattttctt cacttccctt atgcaaaaat ttaccttcag gcccccaaac aatgagaagc
1441 tgagcctgaa gtttagaatg ggtatcacca tttccccagt cagtcaccgc ctctgcgctg
1501 ttcctcaggt gtaatatgt taagaaagaa aggggcaagg aaagtaagaa gacatggcac
1561 gtgttctgaa accactggtg tctgctcaga tgtgttggga caaaatgaaa gtgactttca
1621 agaaagatca gaggaatttg actcagagaa aactagatcc aaatcccagc tctactgtct
1681 cgtccgaatt agccttggga aaatcattta tatgctaaat aatttacctt tttatctagg
1741 agatgaaaag aggataatgt ttccttccat aaagaaagtt cttgtaagaa tcaaaagaa
1801 tgggtgagctt taagtggttt gtaaaccata aaacacatca taaaagttct atctataaaa
1861 aaaaaaaaaa aaaaaa (SEQ ID NO:35)

```

FIGURE 21A

CYP2J2 (cytochrome P450 monooxygenase, NM_000775)

```

LAAMGSLAAALWAVVHPRTL LLGTVAFLLAADFLKRRRPKNYP
PGPWRLPFLGNFFLVDFEQSHLEVQLFVKKYGNLFSLELGDISAVLITGLPLIKEALI
HMDQNFGNRPVTPMREHIFKKNGLIMSSGQAWKEQRRFTLTALRNFGLGKKSLEERIQ
EEAQHLTEAIKEENGQFPDPHFKINNAVSNIICSITFGERFEYQDSWFQQLLKLLDEV
TYLEASKTCQLYNVFPWIMKFLPGPHQTLFSNWKKLKL FVSHMIDKHKRDWNPAETRD
FIDAYLKEMSKHTGNPTSSFHEENLICSTLDLFFAGTETTSTTLRWALLYMALYPEIQ
EKVQAEIDRVIGQQQPSTAARESMFYTNVIVHEVQRMGNI IPLNVPREVTVDTTLAG
YHLPKGMTMLTNTALHRDPTEWATPDTFNPDHFLENGQFKKREAFMPFSIGKRACLG
EQLARTELFIFFTSLMQKFTFRPPNNEKLSLKFRMGITISPVSHRLCAVPQV (SEQ ID

```

NO:36)

FIGURE 21B

31/115

PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214)

```

1  gcccgtctgcg gtaaattgggg cagaggccgg gaggggtggg gggtcccccgc gccgcagcca
61  tggagcagct tcgcgccgcc gcccgctctgc agattgttct gggccacctc ggccgcccct
121 cggccggggc tgtcgtagct catcccactt cagggactat ttccctctgcc agtttccatc
181 ctcaacaatt ccagtatact ctggataata atgttctaac cctggaacag agaaaaatttt
241 atgaagaaaa tgggttttcta gtaatcaaaa atcttgtacc tgatgccgat attcaacgct
301 ttcggaatga gtttgaaaaa atctgcagaa aggaggtgaa accattagga ttaacagtaa
361 tgagagatgt gaccatttcg aaatccgaat atgctccaag tgagaagatg atcacgaagg
421 tccaaggattt ccaggaagat aaggagctct tcagatactg cactctcccc gagattctga
481 aatatgtgga gtgcttcact ggacctaata ttatggccat gcacacaatg ttgataaaca
541 aacctccaga ttctggcaag aagacgtccc gtcacccccct gcaccaggac ctgcactatt
601 tccccttcag gcccagcgat ctcatcgttt gcgcctggac ggcgatggag cacatcagcc
661 ggaacaacgg ctgtctgggt gtgctcccag gcacacacaa gggctccctg aagccccacg
721 attaccccaa gtgggagggg ggaggttaaca aaatgttcca cgggatccag gactacgagg
781 aaaacaaggg ccgggtgcac ctggtgatgg agaagggcga cactgttttc ttccatcctt
841 tgctcatcca cggatctggt cagaataaaa cccagggatt ccggaaggca atttcctgcc
901 atttcgccag tgccgattgc cactacattg acgtgaaggg caccagtcaa gaaaacatcg
961 agaaggaagt tgtaggaata gcacataaat tctttggagc tgaaaatagc gtgaacttga
1021 aggatatttg gatgtttcga gctcgacttg tgaaaggaga aagaaccaat ctttgaaata
1081 gccatctgct ataactcttt caacagaaaa ccaaaaccaa acgaaatgtc taaggaaaat
1141 gttttcttaa tgagatgatg taaccttttc tatcacttgt taaaagcaga aaacatgtat
1201 cagggtactta attgcataga gttagttttg cagcacaatg gtgttgcttt aatggaaaaa
1261 aaaaacagta aaagtgaat attactgttt taaggaaaac taatttaggg tggcagccaa
1321 taaaggtggt tgggtgtctaa ttttaagtgt aaatcaattt ctttcattca gtttagctctt
1381 taccacaagaa gaagtgaatg atttggagct tagggtatgt tttgtatccc ctttctgata
1441 aaccattcc ctaccaattt tatgtcataa gagatttttt tccccaaat ctagaacaat
1501 gtataatata ttcacatcta gtcaagggca taggaacggt gtcatggagt ccaaataaag
1561 tggatattcc tgctcgg (SEQ ID NO:37)

```

FIGURE 22A

PHYH (phytanoyl-CoA-hydroxylase (Refsum disease), NM_006214)

```

MEQLRAAARLQIVLGHILGRPSAGAVVAHPTSGTISSASFHPQQF
QYTLDDNNVLTLEQRKFYEENGFLVIKNLVPDADIQRFRNEFEKICRKEVKPLGLTVMR
DVTISKSEYAPSEKMITKVQDFQEDKELFRYCTLPEILKYVECFTGPNIMAMHTMLIN
KPPDSGKKTSRHPLHQDLHYFFFRPSDLIVCAWTAMEHISRNGCLVVLPGTHKGSILK
PHDYPKWEGBVNKMFHGIQDYEENKARVHLMVEKGDVFFHPLLIHSGQNKTQGRK
AISCHFASADCHYIDVKGTSQENIEKEVVGGIAHKFFGAENSVNLKDIWMFRARLVKGE
RTNL (SEQ ID NO:38)

```

FIGURE 22B

32/115

CYB5 (cytochrome b5, 3' end, NM_001914)

```
1 atggcagagc agtcggacga ggccgtgaag tactacaccc tagaggagat tcagaagcac
61 aaccacagca agagcacctg gctgatcctg caccacaagg tgtacgattt gaccaaattt
121 ctggaagagc atcctggtgg ggaagaagtt ttaagggaaac aagctggagg tgacgctact
181 gagaactttg aggatgtcgg gcactctaca gatgccaggg aaatgtccaa aacattcatc
241 attggggagc tccatccaga tgacagacca aagttaaaca agcctccaga accttaaagg
301 cggtgtttca aggaaactct tatcactact attgattcta gttccagttg gtggaccaac
361 tgggtgatcc ctgccatctc tgcagtggcc gtgccttga tgtatcgctt atacatggca
421 gaggactgaa cacctcctca gaagtcagcg caggaagagc ctgctttgga cacgggagaa
481 aagaagccat tgctaactac ttcaactgac agaaaccttc acttgaaaac aatgatttta
541 atatatctct ttctttttct tccgacatta gaaacaaaac aaaaagaact gtcctttctg
601 cgctcaaatt tttcgagtgt gcctttttat tcactacttt tattttgatg tttccttaat
661 gtgtaattta cttattataa gcatgatctt ttaaaaatat atttggcttt taaagt (SEQ
ID NO:39)
```

FIGURE 23A

CYB5 (cytochrome b5, 3' end, NM_001914)

```
MAEQSDEAVKYYTLEEIQKHNHKSSTWLIILHHKVYDLTKFLEE
PGGEEVLREQAGGDATENFEDVGHSTDAREMSKTFIIGELHPDDRPKLNKPPEP (SEQ ID
NO:40)
```

FIGURE 23B

33/115

COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863)

```
1 cctcctggga gggagctgaa gccgctcgca agactcccgt agtccccacc tctctcagct
61 tccggctggg agtagttccg cttcctgtcc gactgtggtg tctttgctga gggtcacatt
121 gagctgcagg ttgaatccgg ggtgccttta ggattcagca ccatggcgga agacatggag
181 accaaaaatca agaactacaa gaccgcccct tttagacagc gcttcccaa ccagaaccag
241 actagaaact gctggcagaa ctacctggac ttccaccgct gtcagaaggc aatgaccgct
301 aaaggaggcg atatctctgt gtgcgaatgg taccagcgtg tgtaccagtc cctctgcccc
361 acatcctggg tcacagactg ggatgagcaa cgggctgaag gcacgtttcc cgggaagatc
421 tgaactggct gcctctccct ttctctgtgc ctccatcctt ctcccaggat ggtgaagggg
481 gacctggtac ccagtgatcc ccaccccagg atcctaaatc atgacttacc tgctaataaa
541 aactcattgg aaaagtgaaa aaaaaaaaaa aaaaaaaaa (SEQ ID NO:41)
```

FIGURE 24A

COXVIb (coxVIb gene, last exon and flanking sequence, NM_001863)

```
MAEDMETKIKNYKTAPFDSRFPNQTRNCWQNYLDFHRCQKAM
TAKGGDISVCEWYQRVYQSLCPTSWVTDWDEQRAEGTFPGKI (SEQ ID NO:42)
```

FIGURE 24B

34/115

TCF4 (NM_030756)

```

1  gggtttttttt  ttttaccccc  ctttttttatt  tattatttttt  ttgcacattg  agcgggatcct
61  tgggaaacgag  agaaaaaaga  aacccaaact  cacgcgtgca  gaagatctcc  ccccccttcc
121 cctccccctcc  tccctctttt  cccctcccca  ggagaaaaag  acccccaagc  agaaaaaagt
181 tcaccttgga  ctgctctttt  tcttgcaata  ttttttgggg  gggcaaaact  ttgaggggggt
241 gattttttttt  ggctttttctt  cctccttcat  ttttcttcca  aaattgctgc  tgggtgggtga
301 aaaaaaaatg  ccgcagctga  acggcggtgg  aggggatgac  ctaggcgcca  acgacgaact
361 gatttccttc  aaagacgagg  gcgaacagga  ggagaagagc  tccgaaaact  cctcggcaga
421 gagggattta  gctgatgtca  aatcgtctct  agtcaatgaa  tcagaaacga  atcaaaacag
481 ctctctccgat  tccgaggcgg  aaagacggcc  tccgcctcgc  tccgaaagtt  tccgagacaa
541 atccccgggaa  agtttggaag  aagcggccaa  gaggcaagat  ggagggctct  ttaagggggcc
601 accgtatccc  ggctacccct  tcatcatgat  ccccgacctg  acgagccctt  acctcccca
661 cggatcgctc  tcgcccaccg  cccgaacctt  tctccagatg  aaatggccac  tgcttgatgt
721 ccaggcaggg  agcctccaga  gtagacaagc  cctcaaggat  gcccggtccc  catcacgggc
781 acacattgtc  tctaacaaag  tgccagtggg  gcagcaccct  caccatgtcc  accccctcac
841 gcctcttatc  acgtacagca  atgaacactt  cacgccggga  aaccacctc  cacacttacc
901 agccgacgta  gaccccaaaa  caggaatccc  acggcctccg  caccctccag  atatatccc
961 gtattaccca  ctatcgctg  gcaccgtagg  acaaattccc  catccgctag  gatggttagt
1021 accacagcaa  ggtcaaccag  tgtacccaat  cacgacagga  ggattcagac  acccctaccc
1081 cacagctctg  accgtcaatg  cttcctgtgc  caggttcctt  ccccatatgg  tcccaccaca
1141 tcatagctta  cacacgacgg  gcattccgca  tccggccata  gtcacaccaa  cagtcaaaca
1201 ggaatcgctc  cagagtgatg  tcggctcact  ccatagtcca  aagcatcagg  actccaaaaa
1261 ggaagaagaa  aagaagaagc  cccacataaa  gaaacctctt  aatgcattca  tgttgatat
1321 gaaggaaatg  agagcaaagg  tcgtagctga  gtgcacgttg  aaagaaagcg  cggccatcaa
1381 ccagatcctt  gggcgagggt  ggcattgcact  gtccagagaa  gagcaagcga  aatactacga
1441 gctggcccg  aaggagcgac  agcttcatat  gcaactgtac  cccggctggg  ccgcgcggga
1501 taactatgga  aagaagaaga  agaggaagag  ggacaagcag  ccgggagaga  ccaatgaaca
1561 cagcgaatgt  ttcctaaatc  cttgcctttc  acttcctccg  attacagacc  tcagcgctcc
1621 taagaaatgc  cgagcgcgct  ttggccttga  tcaacagaat  aactgggtgc  gcccttgtag
1681 gagaaaaaaa  aagtgcgttc  gctacatata  aggtgaaggc  agctgcctca  gccaccctc
1741 ttcagatgga  agcttactag  attcgcctcc  cccctccccg  aacctgctag  gctccctcc
1801 ccgagacgcc  aagtcacaga  ctgagcagac  ccagcctctg  tcgctgtccc  tgaagcccga
1861 cccctgggcc  cacctgtcca  tgatgcctcc  gccaccgcc  ctctgtctcg  ctgaggccac
1921 ccacaaggcc  tccgccctct  gtcccaacgg  ggccctggac  ctgccccag  ccgctttgca
1981 gcctgccgcc  cctcctcat  caattgcaca  gccgtcgact  tcttggttac  attcccacag
2041 ctccctggcc  gggaccagc  cccagccgct  gtcgctcgtc  accaagtctt  tagaatagct
2101 ttagcgtcgt  gaaccccgct  gctttgttta  tggttttgtt  tcacttttct  taatttgccc
2161 cccaccccca  ccttgaaagg  ttttggtttg  tactctctta  attttggtgc  atgtggctac
2221 attagttgat  gtttatcgag  ttcattgggc  aatatttgac  ccattcttat  ttcaatttct
2281 ctttttaaat  atgtagatga  gagaagaacc  tcatgattgg  taccaaaatt  tttatcaaca
2341 gctgtttaa  gtctttgtag  cgtttaaaaa  atatatatat  atacataact  gttatgtagt
2401 tcggatagct  tagtttttaa  agactgatta  aaaaacaaaa  aaaa (SEQ ID NO:43)

```

FIGURE 25A

35/115

TCF4 (NM_030756)

MPQLNGGGGDDLGANDELISFKDEGEQEEKSSENSSAERDLADV
KSSLVNESETNQNSSSDSEAERRPPRSESFDRDKSRESLEEAQAKRQDGGLFKGPPYPG
YPFIMIPDLTSPYLPNGSLSPARTYLQMKWPLLDVQAGSLQSRQALKDARSPSPAHI
VSNKVPVVQHPHHVHPLTPLITYSNEHFTPGNPPPHLPADVDPKTGIPRPPHPPDISP
YYPLSPGTVGQIPHPLGWLVPQQGQPVYPIITGGFRHPYPTALTVNASVSRFPPHMVP
PHHTLHTTGIPHPAIVTPTVKQESSQSDVGSLSHSSKHQDSKKEEEKKKPHIKKPLNAF
MLYMKEMRAKVVAECTLKESAAINQILGRRWHALSREEQAKYYELARKERQLHMQLYP
GWSARDNYGKKKKRKRDKQPGETNEHSECFLNPCLSLPPIITDLSAPKKCRARFGLDQQ
NNWCGPCRRKKKCVRYIQGEGSCLSPSSDGSLLDSPPSPNLLGSPPRDAKSQTEQT
QPLSLSLKPDPLAHLMMPPPPALLAEATHKASALCPNGALDLPPAALQPAAPSSSI
AQPSTSWLHSHSSLAGTQPQPLSLVTKSLE (SEQ ID NO:44)

FIGURE 25B

36/115

CAD17 (liver-intestine cadherin, NM_004063)

```

1 agggagtggt cccgggggag atactccagt cgtagcaaga gtctcgacca ctgaatggaa
61 gaaaaggact ttttaaccacc atttttgtgac ttacagaaaag gaatttgaat aaagaaaact
121 atgatacttc aggcccatct tcaactccctg tgtcttctta tgctttatatt ggcaactgga
181 tatggccaag aggggaagtt tagtggaccc ctgaaaccca tgacattttc tatttatgaa
241 ggccaagaac cgagtcaaata tatattccag ttttaaggcca atoctoctgc tgtgactttt
301 gaactaactg gggagacaga caacatattt gtgatagaac gggagggact tctgtattac
361 aacagagcct tggacaggga aacaagatct actcacaatc tccagggttg agccctggac
421 gctaattgaa ttatagtgga ggggtccagtc cctatcacca tagaagtga ggacatcaac
481 gacaatcgac ccacgtttct ccagtcaaag tacgaaggct cagtaaggca gaactctcgc
541 ccaggaaagc ccttcttgta tgtcaatgcc acagacctgg atgatccggc cactcccaat
601 ggccagcttt attaccagat tgtcatccag cttcccatga tcaacaatgt catgtacttt
661 cagatcaaca acaaaacggg agccatctct cttacccgag agggatctca ggaattgaat
721 cctgctaaga atccttecta taatctggtg atctcagtga aggacatggg aggccagagt
781 gagaattcct tcagtgtatc cacatctgtg gatatcatag tgacagagaa tatttggaaa
841 gcacaaaaac ctgtggagat ggtggaaaac tcaactgac ctcaccccat caaaatcact
901 caggtgcggt ggaatgatcc cgggtgcacaa tattccttag ttgacaaaga gaagctgcc
961 agattcccat tttcaattga ccaggaagga gatatttacg tgactcagcc cttggaccga
1021 gaagaaaagg atgcatatgt tttttatgca gttgcaaagg atgagtacgg aaaaccactt
1081 tcataatccg tggaaattca tgtaaaagtt aaagatatta atgataatcc acctacatgt
1141 ccgtcaccag taaccgtatt tgaggtccag gagaatgaac gactgggtaa cagtatcggg
1201 acccttactg cacatgcacg ggatgaagaa aatactgcca acagttttct aaactacagg
1261 attgtggagc aaactcccaa acttcccatg gatggactct tcctaattca aacctatgct
1321 ggaatgttac agttagctaa acagtccttg aagaagcaag atactcctca gtacaactta
1381 acgatagagg tgtctgacaa agatttcaag accctttgtt ttgtgcaaat caacgttatt
1441 gatatcaatg atcagatccc catctttgaa aaatcagatt atggaaacct gactcttgct
1501 gaagacacaa acattgggtc caccatctta accatccagg ccatgatgc tgatgagcca
1561 tttactggga gttctaaaat tctgtatcat atcataaagg gagacagtga gggacgcctg
1621 ggggttgaca cagatcccca taccaacacc ggatatgtca taattaaaaa gcctcttgat
1681 tttgaaacag cagctgtttc caacattgtg ttcaaagcag aaaatcctga gcctctagt
1741 tttggtgtga agtacaatgc aagttctttt gccaaagtca cgcttattgt gacagatgtg
1801 aatgaagcac ctcaattttc ccaacacgta ttccaagcga aagtcagtga ggatgtagct
1861 ataggcacta aagtgggcaa tgtgactgcc aaggatccag aaggtctgga cataagctat
1921 tcaactgagg gagacacaag aggttggtctt aaaattgacc acgtgactgg tgagatcttt
1981 agtgtggctc cattggacag agaagccgga agtccatctc gggtaacaag ggtggccaca
2041 gaagtggggg ggtcttcctt gagctctgtg tcagagttcc acctgatcct tatggatgtg
2101 aatgacaacc ctcccaggct agccaaggac tacacgggct tgttcttctg ccatcccctc
2161 agtgcacctg gaagtctcat tttcgaggct actgatgatg atcagcactt atttcggggg
2221 cccattttta cattttccct cggcagtggg agcttacaaa acgactggga agtttccaaa
2281 atcaatggta ctcatgccg actgtctacc aggcacacag agtttgagga gagggagtat
2341 gtogtcttga tccgcatcaa tgatgggggt cggccaccct tggaggcat tgttcttta
2401 ccagttacat tctgcagttg tgtggaagga agttgtttcc ggccagcagg tcaccagact
2461 gggataccca ctgtgggcat ggcagttggt atactgctga ccacccttct ggtgattggt
2521 ataattttag cagttgtgtt tatccgcata aagaaggata aaggcaaaga taatgttgaa
2581 agtgctcaag catctgaagt caaacctctg agaagctgaa tttgaaaagg aatgtttgaa
2641 tttatatagc aagtgttatt tcagcaacaa ccatctcatc ctattacttt tcatctaacg
2701 tgcattataa ttttttaaac agatattccc tcttgtcctt taatatttgc taaatatttc
2761 ttttttgagg tggagtcttg ctctgtcgcc caggctggag tacagtgggt tgatcccagc
2821 tgggtgcaac ctccgctcc tgggttcaca tgattctcct gcctcagctt cctaagtagc
2881 tcgggtttaca ggcaccacc accatgcoca gctaattttt gtatttttaa tagagacggg
2941 gtttcgccat ttggccaggc tgggtcttgaa ctctgacgt caagtgatct gcctgccttg
3001 gtctcccaat acaggcatga accactgcac ccacctactt agatatttca tgtgtatag
3061 acattagaga gatttttcat ttttccatga catttttctt ctctgcaaat ggcttagcta

```

FIGURE 26A

37/115

```
3121 cttgtgtttt tcccttttgg ggcaagacag actcattaaa tattctgtac attttttctt
3181 tatcaaggag atatatcagt gttgtctcat agaactgcct ggattccatt tatgtttttt
3241 ctgattccat cctgtgtccc cttcatcctt gactcctttg gtatttcact gaatttcaaa
3301 catttgtcag agaagaaaaa cgtgaggact caggaaaaat aaataaataa aagaacagcc
3361 ttttccctta gtattaacag aaatgtttct gtgtcattaa ccatctttaa tcaatgtgac
3421 atgttgctct ttggctgaaa ttcttcaact tggaaatgac acagaccac agagggtgtt
3481 caaacacaac ctactctgca aaccttggtt aaggaaaccag tcagctggcc agatttcctc
3541 actacctgcc atgcatacat gctgcgcacg ttttcttcat tcgtatgtta gtaaagtttt
3601 ggttattata tatttaacat gtggaagaaa acaagacatg aaaagagtgg tgacaaatca
3661 agaataaaca ctggttgtag tcagttttgt ttgttaa (SEQ ID No:45)
```

FIGURE 26B

38/115

CAD17 (liver-intestine cadherin, NM_004063)

MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFESIYEGQEP
SQIIFQFKANPPAVTFELTGETDNIFVIEREGLLYNRLDRETRSTHNLQVAALDAN
GIIVEGPVPITIEVKDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPN
GQLYYQIVIQLPMINNVMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGG
QSENSFSDTTSVDIIVTENIWKAPKPVEMVENSTDPHPKITQVRWNDPGAQYSLVDK
EKLPRFPFSIDQEGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDIN
DNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDG
LFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKDFKTLCFVQINVIDINDQIPIFE
KSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHI IKGDSEGR LGVDTPHT
NTGYVI IKKPLDFETA AVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQF
SQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAP
LDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTG LFFCHPLSA
PGSLIFEATDDDQHLFRGPHTFSLGSGSLQNDWEVSKINGTHARLSTRHTEFEEREY
VVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQGTGIPTVGMVAVGILLTTLLV
IGIILAVVFIRIKKDKGKDNVESQAQASEVKPLRS (SEQ ID NO:46)

FIGURE 26C

39/115

CLDN15 (claudin 15, NM_014343)

```

1  ctcgtcaaca gctgccgcgc gcaggcttag ctcattcctc tgacctgcca ggaagcagag
61 agacccacag agcaggaggg aggcagaaag tggagacgga cctgagcccg aggaagaggc
121 aggcagaggc tgaggctgat tccaccccag cctgcctgga caaccctcct tagccgcagc
181 cccttccagt tccctagggg ttctgcccct cccctctctt ggggcaccag ccccccaggg
241 tccctgcatcc caccatgtcg atggctgtgg aaacctttgg cttcttcatg gcaactgtgg
301 ggctgctgat gctgggggtg actctgccaa acagctactg gcgagtgtcc actgtgcacg
361 ggaacgtcat caccaccaac accatcttcg agaacctctg gtttagctgt gccaccgact
421 ccctgggcgt ctacaactgc tgggagttcc cgtccatgct ggccctctct gggtatattc
481 aggcctgccg ggcactcatg atcacgcgca tccctcctggg cttcctcggc ctcttgctag
541 gcatagcggg cctgcgctgc accaacattg ggggcctgga gctctccagg aaagccaagc
601 tggcggccac cgcaggggcc ctccacattc tggccggtat ctgcgggatg gtggccatct
661 cctggtagcg cttcaacatc acccgggact tcttcgaccc cttgtacccc ggaaccaagt
721 acgagctggg ccccgccctc tacctggggt ggagcgctc actgatctcc atcctgggtg
781 gcctctgcct ctgctccgcc tgctgctgog gctctgacga ggaccagcc gccagcgccc
841 ggcggcccta ccaggctccc gtgtccgtga tgcccgctcg cacctcggac caagaaggcg
901 acagcagctt tggcaaatac ggcagaaacg cctacgtgta gcagctctgg cccgtggggc
961 ccgctgtctt cccactgccc caaggagagg ggacctggcc gggggccatt cccctatagt
1021 aacctcaggg gccggccacg ccccgctccc gtagccccgc cccggccacg gccccgtgtc
1081 ttgcaactct atggcccctc caggccaaga actgctcttg ggaagtgcga tatctccct
1141 ctgaggctgg atccctcatc ttctgacctt gggttctggg ctgtgaaggg gacggtgtcc
1201 ccgcacgttt gtattgtgta taaatacatt cattaataaa tgcattattg gaccgttc

```

(SEQ ID NO:47)

FIGURE 27A

CLDN15 (claudin 15, NM_014343)

```

MSMAVETFGFFMATVGLLMLGVTL PNSYWRVSTVHGNVITNTI
FENLWFSCATDSLGVYNCWEFPSMLALSGYIQACRALMITAILLGFLLGLGIAGLRC
TNIGGLELSRKAKLAATAGALHILAGICGMVAISWYAFNITRDFFDPLYPGTKYELGP
ALYLGSASLISILGGLCLCSACCCGSDDEPAASARRPYQAPVSVMPVATSDQEGDSS
FGKYGRNAYV (SEQ ID NO:48)

```

FIGURE 27B

40/115

CFTR (chloride channel, NM_000492)

```

1  aattggaagc aaatgacatc acagcaggtc agagaaaaag ggttgagcgg caggcaccca
61  gagtagtagg tctttggcat taggagcttg agcccagacg gccctagcag ggaccccagc
121 gcccgagaga ccatgcagag gtcgcctctg gaaaaggcca gcgttgcttc caaacttttt
181 ttcagctgga ccagaccaat tttgaggaaa ggatacagac agcgcctgga attgtcagac
241 atataccaaa tcccttctgt tgattctgct gacaatctat ctgaaaaatt ggaaagagaa
301 tgggatatag agctggcttc aaagaaaaat cctaaactca ttaatgccct tcggcgatgt
361 tttttctgga gatttatgtt ctatggaatc tttttatatt taggggaagt caccaaagca
421 gtacagcctc tcttactggg aagaatcata gcttcctatg acccgataa caaggaggaa
481 cgctctatcg cgatttatct aggcataggc ttatgccttc tctttattgt gaggacactg
541 ctctacacc cagccatttt tggccttcac cacattggaa tgcagatgag aatagctatg
601 tttagtttga tttataagaa gactttaaag ctgtcaagcc gtgttctaga taaaataagt
661 attggacaac ttgttagtct cttttccaac aacctgaaca aatttgatga aggacttgca
721 ttggcacatt tcgtgtggat cgctcctttg caagtggcac tcctcatggg gctaactctgg
781 gagttgttac aggcgtctgc cttctgtgga ctctggttcc tctggtttcc tgcccttttt
841 caggctgggc tagggagaat gatgatgaag tacagagatc agagagctgg gaagatcagt
901 gaaagacttg tgattacctc agaaatgatt gaaaatatcc aatctgttaa ggcatactgc
961 tgggaagaag caatggaaaa aatgattgaa aacttaagac aaacagaact gaaactgact
1021 cgggaaggcag cctatgtgag atacttcaat agctcagcct tcttcttctc agggttcctt
1081 gtggtgtttt tatctgtgct tccctatgca ctaatcaaag gaatcatcct ccggaaaata
1141 ttcaccacca tctcattctg cattgttctg cgcattggcg tcaactcgga atttccctgg
1201 gctgtacaaa catggtatga ctctcttgga gcaataaaca aaatacagga tttcttacia
1261 aagcaagaat ataagacatt ggaatataac ttaacgacta cagaagtagt tagggagaa
1321 gtaacagcct tctgggagga gggatttggg gaattatttg agaaagcaaa acaaaacaat
1381 aacaatagaa aaacttctaa tggatgatgc agcctcttct tcagtaattt ctcacttctt
1441 ggtactcctg tctgaaaga tattaatttc aagatagaaa gaggacagtt gttggcggtt
1501 gctggatcca ctggagcagg caagacttca cttctaata tgattatggg agaactggag
1561 ccttcagagg gtaaaattaa gcacagtggg agaatttcat tctgttctca gttttcctgg
1621 attatgcctg gcaccattaa agaaaatatc atctttggtg tttcctatga tgaatataga
1681 tacagaagcg tcatcaaagc atgccaacta gaagaggaca tctccaagtt tgcagagaaa
1741 gacaatatag ttcttggaga aggtggaatc acactgagtg gaggtcaacg agcaagaatt
1801 tcttttagcaa gagcagtata caaagatgct gatttgatct tattagactc tccttttggg
1861 tacctagatg ttttaacaga aaaagaaata tttgaaagct gtgtctgtaa actgatggct
1921 aacaaaacta ggattttggt cacttctaaa atggaacatt taaagaaagc tgacaaaata
1981 ttaattttga atgaaggtag cagctatttt tatgggacat tttcagaact ccaaaatcta
2041 cagccagact ttagctcaaa actcatggga tgtgattctt tcgaccaatt tagtgcagaa
2101 agaagaaatt caatcctaac tgagacctta caccgtttct cattagaagg agatgctcct
2161 gtctcctgga cagaaacaaa aaaacaatct tttaaacaga ctggagagtt tggggaaaaa
2221 aggaagaatt ctattctcaa tccaatcaac tctatacgaa aattttccat tgtgcaaaag
2281 actcccttac aaatgaatgg catcgaagag gattctgatg agcctttaga gagaaggctg
2341 tccttagtac cagattctga gcaggagag gcgatactgc ctcgcacag cgtgatcagc
2401 actggcccca cgcttcaggc acgaaggagg cagtctgtcc tgaacctgat gacacactca
2461 gttaaaccaag gtcagaacat tcaccgaaag acaacagcat ccacacgaaa agtgtcactg
2521 gccctcagg caaacttgac tgaactggat atatatcaa gaaggttatc tcaagaaact
2581 ggcttggaat taagtgaaga aattaacgaa gaagacttaa aggagtgcct ttttgatgat
2641 atggagagca taccagcagt gactacatgg aacacatacc ttcgatatat tactgtccac
2701 aagagcttaa tttttgtgct aatttggtgc ttagtaattt ttctggcaga ggtggctgct
2761 tctttgggtg tgctgtggct ccttggaaac actcctcttc aagacaaagg gaatagtact
2821 catagtagaa ataacagcta tgcagtgatt atcaccagca ccagttcgta ttatgtgttt
2881 tacatttacg tgggagtagc cgacactttg cttgctatgg gattcttcag aggtctacca
2941 ctggtgcata ctctaaccac agtgcgaaa attttacacc acaaaatggt acattctgtt
3001 cttcaagcac ctatgtcaac cctcaacagc ttgaaagcag gtgggattct taatagattc
3061 tccaaagata tagcaatttt ggatgacctt ctgcctctta ccataattga cttcatccag

```

FIGURE 28A

41/115

```

3121 ttgttattaa ttgtgattgg agctatagca gttgtgcgag ttttacaacc ctacatcttt
3181 gttgcaacag tgccagtgat agtggctttt attatgttga gagcatattt cctccaaacc
3241 tcacagcaac tcaaacaact ggaatctgaa ggcaggagtc caattttcac tcatcttggt
3301 acaagcttaa aaggactatg gacacttcgt gccttcggac ggcagcctta ctttgaaact
3361 ctgttccaca aagctctgaa ttacataact gccaaactggg tcttgtacct gtcaacactg
3421 cgctggttcc aaatgagaat agaaatgatt tttgtcatct tcttcattgc tgttaccttc
3481 atttccattt taacaacagg agaaggagaa ggaagagttg gtattatcct gactttagcc
3541 atgaatatca tgagtacatt gcagtgggct gtaaactcca gcatagatgt ggatagcttg
3601 atgcgatctg tgagccgagt ctttaagttc attgacatgc caacagaagg taaacctacc
3661 aagtcaacca aaccatacaa gaatggccaa ctctcgaaag ttatgattat tgagaattca
3721 cacgtgaaga aagatgacat ctggccctca gggggccaaa tgactgtcaa agatctcaca
3781 gcaaaatata cagaaggtgg aaatgccata ttagagaaca tttccttctc aataagtcct
3841 ggccagaggg tgggcctctt ggaagaact ggatcagggg agagtacttt gttatcagct
3901 tttttgagac tactgaacac tgaaggagaa atccagatcg atggtgtgtc ttgggattca
3961 ataactttgc aacagtggag gaaagccttt ggagtgtacac cacagaaagt atttattttt
4021 tctggaacat ttagaaaaaa cttggatccc tatgaacagt ggagtgtaca agaatatgg
4081 aaagttgcag atgaggttgg gctcagatct gtgatagaac agtttcctgg gaagcttgac
4141 tttgtccttg tggatggggg ctgtgtccta agccatggcc acaagcagtt gatgtgcttg
4201 gctagatctg ttctcagtaa ggcgaagatc ttgctgcttg atgaaccag tgctcatttg
4261 gatccagtaa cataccaaata aattagaaga actctaaaac aagcatttgc tgattgcaca
4321 gtaattctct gtgaacacag gatagaagca atgctggaat gccacaatt tttggtcata
4381 gaagagaaca aagtgcggca gtacgattcc atccagaaac tgctgaacga gaggagcctc
4441 ttccggcaag ccacagccc ctccgacagg gtgaagctct tccccaccg gaactcaagc
4501 aagtgcaggt ctaagcccca gattgctgct ctgaaagagg agacagaaga agaggtgcaa
4561 gatacaaggc tttagagagc agcataaatg ttgacatggg acatttgctc atggaattgg
4621 agctcgtggg acagtcacct catggaattg gagctcgtgg aacagttacc tctgcctcag
4681 aaaacaagga tgaattaagt ttttttttaa aaaagaaaca tttggtaagg ggaattgagg
4741 acactgatat gggctcttgat aaatggcttc ctggcaatag tcaaattgtg tgaaaggtac
4801 ttcaaactct tgaagattta ccacttgtgt tttgcaagcc agattttcct gaaaaccctt
4861 gccatgtgct agtaattgga aaggcagctc taaatgtcaa tcagcctagt tgatcagctt
4921 attgtctagt gaaactcgtt aatttgtagt gttggagaag aactgaaatc atacttctta
4981 ggggttatgat taagtaatga taactggaaa cttcagcggg ttatataagc ttgtattcct
5041 ttttctctcc tctcccatg atgtttagaa acacaactat attgtttgct aagcattcca
5101 actatctcat ttccaagcaa gtattagaat accacaggaa ccacaagact gcacatcaaa
5161 atatgccccca ttcaacatct agtgagcagt caggaaaagag aacttccaga tcctggaaat
5221 caggggttagt attgtccagg tctacaaaaa atctcaatat ttcagataat cacaatacat
5281 cccttacctg ggaaagggct gttataatct ttcacagggg acaggatggg tccttgatg
5341 aagaagtga tatgcctttt cccaactcca gaaagtgaca agctcacaga cctttgaact
5401 agagtttagc tggaaaagta tgttagtgca aattgtcaca ggacagccct tctttccaca
5461 gaagctccag gttagaggtg tgtaagtaga taggccatgg gcactgtggg tagacacaca
5521 tgaagtccaa gcatttagat gtataggttg atggtggtat gttttcaggc tagatgtatg
5581 tacttcatgc tgtctacact aagagagaat gagagacaca ctgaagaagc accaatcatg
5641 aattagtttt atatgcttct gttttataat tttgtgaagc aaaatTTTTT ctctaggaaa
5701 tatttatTTT aataatgttt caaacatata ttacaatgct gtattttaaa agaattgatta
5761 tgaattacat ttgtataaaa taatttttat atttgaaata ttgacttttt atggcaactag
5821 tatttttatg aaatattatg ttaaaactgg gacaggggag aacctagggt gatattaacc
5881 aggggccatg aatcaccttt tggctcggag ggaagccttg gggctgatcg agttgtgccc
5941 cacagctgta tgattcccag ccagacagag cctcttagat gcaagtgggt agaagattgg
6001 accaccagtc tgactgtttc catcaagggt acactgcctt ctcaactcca aactgactct
6061 taagaagact gcattatatt tattactgta agaaaatat acttgtcaat aaaatccata
6121 catttgtgt (SEQ ID NO:49)

```

FIGURE 28B

42/115

CFTR (chloride channel, NM_000492)

MQRSPLEKASVVSKLFFSWTRPILRKGYRQRLELSDIYQIPSDV
SADNLSEKLEREWDRELASKKNPKLINALRRCFFWRFMFYGIFLYLGEVTKAVQPLLL
GRIIASYDPDNKEERSIAIYLGIGLCCLLFIVRTLLLLHPAIFGLHHIGMQMRIAMFSLI
YKKTLLKSSRVLDKISIGQLVSLLSNNLNKFDDEGLALAHFVWIAPLQVALLMGLIWEL
LQASAFCGLGFLIVLALFQAGLGRMMMKYRDQRAGKISERLVITSEMIEINIQSVKAYC
WEEAMEKMIENLRQTELKLRKAAYVRYFNSSAFFFSGFFVVFSLVLPYALIKGIILR
KIFTTISFCIVLRMAVTRQFPWAVQTWYDSLGAINKIQDFLQKQEKTYEYNLT'TTEV
MENVTAFWEEGFGELFEKAKQNNNNRKT'SNGDDSLFFSNFSLLGTPVLKDINFKIER
QLLAVAGSTGAGKTSLLMMIMGELEPSEGKIKHSGRISFCSQFSWIMPGTIKENIIF
VSYDEYRYSVIKACQLEEDISKFAEKDNIVLGEKGITLSGGQRARISLARAVYKDA
LYLLDSPFGYLDVLTEKEIFESCVCCKLMANKTRILVTSKMEHLKKADKILILNEGSS
FYGTFSELQNLQPDFSSKLMGCDSFDQFSAERRNSILTETLHRFSLEGDAPVSWTET
KQSFKQTGEFGEKRKNSILNPINSIRKFSIVQKTPLQMNIEEDSDEPLERRLSLVP
SEQGEAILPRISVISTGPTLQARRRQSVLNLMTHSVNOGQNIHRKT'TASTRKVSLAP
ANLTEDIYSRRLSQETGLEISEEINEEDLKECLFDDMESIPAVTTWNTYLYRYITVH
SLIFVLIWCLVIFLAEVAASLVVLWLLGNTPLQDKGNSTHSRNNSYAVIITSTSSYY
FYIYVGVAADTLLAMGFFRGLPLVHTLITVSKILHHKMLHSLVQAPMSTLNTLKAGGI
NRFSKDIAILDLLPLTIFDFIQLLLIVIGAIAVVAVLQPYIFVATVPVIVAFIMLR
YFLQTSQQLKQLESEGRSPIFTHLVTSKGLWTLRAFGROPYFETLFHKALNLHTAN
FLYLSTLRWFQMRIEMI FVIFFI AVTFISILT TGEGEGRVGI I LTLAMNIMSTLQWA
NSSIDVDSLMRSVSRVFKFIDMPTEGKPTKSTKPYKNGQLSKVMI I ENSHVKKDDIW
SGGQMTVKDLTAKYTEGGNAILENISFSISPGRVGLLGRTGSGKSTLLSAFLRLLN
EGEIQIDGVSWDSITLQQWRKAFGVIPQKVFI FSGTFRKNLDPYEQWSDQEIWKVAD
VGLRSVIEQFPKLD FVLVDGGCVLSHGKQLMCLARSVLSKAKILLLDEPSAHLDP
TYQII RRTLKQAFADCTVILCEHRIEAMLECCQFLVIEENKVRQYDSIQKLLNERSL
RQAISPSDRVKLFFHRNSSKCKSKPQIAALKEETEEVEVQDTRL (SEQ ID NO:50)

FIGURE 28C

43/115

H2R (histamine H2 receptor, NM_022304)

```

1  ctccctgccct ccactgactc cagagagggga gatccccagt acttgactcc atcacgcaga
61  tgggagcagg caccagctat ggagagggat acagctgcgt ctccacatga cccatcctgc
121 atgacaccaa agccaccgcc agacagtgcc tgggattcta tgcaaaacct gggaagcgga
181 gacctacccc agccccggga ggaagctagc tcttcagggg accgtctgag gactggagtt
241 tgatccatga acctggcttc gaggccttgc tttctctct tcttcattca tattcattcc
301 caacacctta gaagggtgtg cttaatttat ttctagaaaa gcagcccaga gtcagtcatt
361 gaagccttcc ccaccccctg gccaaaaaaa aaaaaaaaaa aaaactggac acattttgga
421 tctgttggga gcttggagtc cagtggttgg catagtgtgc acattgggag cagagaagaa
481 gcaaccaggg gccctgatca ggggactgag ccgtagagtc ccaggatggc acccaatggc
541 acagcctctt ctttttgcc ggactctacc gcatgcaaga tcaccatcac cgtggctcctt
601 gcggctcctca tcctcatcac cgttgcctggc aatgtggtcg tctgtctggc cgtgggcttg
661 aaccgccggc tccgcaacct gaccaattgt ttcctcgtgt ccttggctat cactgacctg
721 ctccctggcc tcctgggtgt gcccttctct gccatctacc agctgtcctg caagtggagc
781 tttggcaagg tcttctgcaa tatctacacc agcctggatg tgatgctctg cacagcctcc
841 attcttaacc tcttcatgat cagcctcgac cggtaactgc ctgtcatgga cccactgcgg
901 taccctgtgc tggtcacccc agttcgggtc gccatctctc tggctttaat ttgggtcatc
961 tccattaccc tgtcctttct gtctatccac ctgggggtga acagcaggaa cgagaccagc
1021 aagggcaatc ataccacctc taagtgcata gtccagggtca atgaagtgtg cgggctgggtg
1081 gatgggctgg tcaccttcta cctcccgcta ctgatcatgt gcatcaccta ctaccgcac
1141 ttcaagggtcg cccgggatca ggccaagagg atcaatcaca ttagctcctg gaaggcagcc
1201 accatcaggg agcacaagc cacagtgaca ctggccgccg tcatgggggc cttcatcatc
1261 tgctggtttc cctacttcac cgcgtttgtg taccgtgggc tgagagggga tgatgccatc
1321 aatgaggtgt tagaagccat cgttctgtgg ctgggctatg ccaactcagc cctgaacccc
1381 atcctgtatg ctgcgctgaa cagagacttc cgcaccgggt accaacagct cttctgctgc
1441 aggctggcca accgcaactc ccacaaaact tctctgaggt ccaacgcctc tcagctgtcc
1501 aggacccaaa gccgagaacc caggcaacag gaagagaaac cctgaagct ccaggtgtgg
1561 agtgggacag aagtcacggc cccccaggga gccacagaca ggtaatagcc ctagccattg
1621 gtgcacagga tgggggcaat gggaggggat gctactgatg ggaatgatta agggagctgc
1681 tgttttaggtg gtgctggttt atgttctagg aactcttcat gagcactttg taaacaccct
1741 cttgcttaat cctcccaacg gccccaaaag gtagaactta gtcctctttt aaaaggagca
1801 cattaaaatt ctcagaggac ttggcaaggg ccgcacagct ggggcat (SEQ ID NO:51)

```

FIGURE 29A

H2R (histamine H2 receptor, NM_022304)

```

APNGTASSFCLDSTACKITITVVLAVLILITVAGNVVCLAVG
NRRLRNLTNCFIVSLAITDLLLGLLVLPFSAIYQLSCKWSFGKVFCNIYTSLDVMLC
ASILNLFMISLDRYCAVMDPLRYPVLVTPVRVAISLVLIWVISITLSFLSIHLGWNS
NETSKGNHTTSKCKVQVNEVYGLVDGLVTFYLPPLLIMCITYYRIFKVARDAQKRINH
SSWKAATIREHKATVTLAAVMGAFIICWFPYFTAFVYRGLRGDDAINEVLEAIVLWL
YANSALNPILYAALNRDFTGYQQLFCCRLANRNSHKTSLSRSNASQLSRTQSREPRQ
EEKPLKLQVWSGTEVTAPQGATDR (SEQ ID NO:52)

```

FIGURE 29B

44/115

EGFR (NM_005228)

```

1 gagctagccc cggcggccgc cgcgcgccag accggacgac aggccacctc gtcggcgctcc
61 gcccagagtcc ccgcctcgcc gccaacgccca caaccaccgc gcacggcccc ctgactccgt
121 ccagtatattga tcgggagagc cggagcgagc tcttcgggga gcagcgatgc gaccctccgg
181 gacggccggg gcagcgctcc tggcgctgct ggctgcgctc tgcccggcga gtcgggctct
241 ggaggaaaag aaagtttgcc aaggcacgag taacaagctc acgcagttgg gcacttttga
301 agatcatttt ctcagcctcc agaggatgtt caataactgt gaggtgggtcc ttgggaattt
361 ggaaattacc tatgtgcaga ggaattatga tctttccttc ttaaagacca tccaggaggt
421 ggctgggttat gtccctcattg ccctcaacac agtggagcga attcctttgg aaaacctgca
481 gatcatcaga ggaaatatgt actacgaaaa ttcctatgcc ttagcagctc tatctaacta
541 tgatgcaaat aaaaccggac tgaaggagct gcccatgaga aatttacagg aaatcctgca
601 tggcgcgctg cggttcagca acaaccctgc cctgtgcaac gtggagagca tccagtggcg
661 ggacatagtc agcagtgact ttctcagcaa catgtcgatg gacttcagca accacctggg
721 cagctgccaa aagtgtgatc caagctgtcc caatgggagc tgctggggtg caggagagga
781 gaactgccag aaactgacca aaatcatctg tgcccagcag tgctccgggc gctgccgtgg
841 caagtcccc agtgactgct gccacaacca gtgtgctgca ggctgcacag gcccccgga
901 gagcgactgc ctggctctgcc gcaaattccg agacgaagcc acgtgcaagg acacctgccc
961 cccactcatg ctctacaacc ccaccacgta ccagatggat gtgaaccccc agggcaataa
1021 cagctttggg gccacctgcy tgaagaagtg tccccgtaat tatgtggtga cagatcacgg
1081 ctctgctgct cgagcctgtg gggccgacag ctatgagatg gaggaagacg gcgtccgcaa
1141 gtgtaagaag tgcgaagggc cttgccgcaa agtgtgtaac ggaataggta ttggtgaatt
1201 taaagactca ctctccataa atgtacgaa tattaacac ttcaaaaact gcacctccat
1261 cagtggcgat ctccacatcc tgccggtggc atttaggggt gactccttca cacatactcc
1321 tcctctggat ccacaggaac tggatattct gaaaaccgta aaggaaatca cagggttttt
1381 gctgattcag gcttggcctg aaaacaggac ggacctccat gcctttgaga acctagaaat
1441 catacgcgcc aggaccaagc aacatggtca gttttctctt gcagtcgtca gctgaacat
1501 aacatccttg ggattacgct ccctcaagga gataagtgat ggagatgtga taatttcagg
1561 aaacaaaaat ttgtgctatg caaatacaat aaactggaaa aaactgttg ggagctccg
1621 tcagaaaacc aaaattataa gcaacagagg tgaaaacagc tgcaaggcca caggccaggt
1681 ctgccatgcc ttgtgctccc ccgagggtctg ctggggcccg gagccagggt actgcgtctc
1741 ttgccggaat gtcagccgag gcagggaatg cgtggacaag tgcaaccttc tggagggtga
1801 gccaaaggag tttgtggaga actctgagtg catacagtg caccagagt gcctgcctca
1861 ggcatgaac atcacctgca caggacgggg accagacaac tgtatccagt gtgcccacta
1921 cattgacggc cccactgcy tcaagacctg ccggcagga gtcatgggag aaaacaacac
1981 cctggtctgg aagtacgcag acgcggcca tgtgtgccac ctgtgccatc caaactgcac
2041 ctacggatgc actgggccag gtcttgaagg ctgtccaacg aatgggccta agatcccgtc
2101 catcgccact gggatggtgg gggccctcct ctgtctgctg gtggtggccc tggggatcgg
2161 cctcttcattg cgaaggcgcc acatcgttcg gaagcgcacg ctgctggaggc tgctgcagga
2221 gagggagctt gtggagcctc ttacaccagc tggagaagct cccaaccaag ctctcttgag
2281 gatcttgaag gaaactgaat tcaaaaagat caaagtgtg ggctccggtg cgttcggcac
2341 ggtgtataag ggactctgga tcccagaagg tgagaaagtt aaaattccc tgctatcaa
2401 ggaattaaga gaagcaacat ctccgaaagc caacaaggaa atcctcgatg aagcctacgt
2461 gatggccagc gtggacaacc cccacgtgtg ccgctgctg ggcactctgc tcacctccac
2521 cgtgcagctc atcacgcagc tcatgccctt cggctgcctc ctggactatg tccgggaaca
2581 caaagacaat attggctccc agtacctgt caactggtgt gtgcagatcg caaaggcat
2641 gaactacttg gaggaccgtc gtctggtgca ccgcgacctg gcagcagga acgtactgt
2701 gaaaacaccg cagcatgtca agatcacaga ttttgggctg gccaaactgc tgggtgcgga
2761 agagaaagaa taccatgcag aaggaggcaa agtgctatc aagtggatgg cattggaatc
2821 aattttacac agaattctata cccaccagag tgatgtctgg agctacgggg tgaccgtttg
2881 ggagttgatg acctttggat ccaagccata tgacggaatc cctgccagcg agatctcctc
2941 catcctggag aaaggagaac gcctccctca gccaccata tgtaccatcg atgtctacat
3001 gatcatggtc aagtgtgga tgatagacgc agatagtcgc ccaaagttcc gtgagttgat
3061 catcgaattc tccaaaatgg cccgagaccc ccagcgctac cttgtcattc agggggatga

```

FIGURE 30A

45/115

```

3121 aagaatgcat ttgccaagtc ctacagactc caactttctac cgtgccctga tggatgaaga
3181 agacatggac gacgtggtgg atgccgacga gtacctcatc ccacagcagg gcttcttcag
3241 cagccccctcc acgtcacgga ctccccctcc gagctctctg agtgcaacca gcaacaattc
3301 caccgtggct tgcattgata gaaatgggct gcaaagctgt cccatcaagg aagacagctt
3361 cttgcagcga tacagctcag accccacagg cgcccttgact gaggacagca tagacgacac
3421 cttcctccca gtgcctgaat acataaacca gtccgttccc aaaaggcccc ctggctctgt
3481 gcagaatcct gtctatcaca atcagcctct gaaccccgcg ccagcagag acccacacta
3541 ccaggacccc cacagcactg cagtgggcaa ccccgagtat ctcaacactg tccagcccac
3601 ctgtgtcaac agcacattcg acagccctgc ccactggggc cagaaaggca gccaccaaat
3661 tagcctggac aaccctgact accagcagga cttctttccc aaggaagcca agccaaatgg
3721 catctttaag ggctccacag ctgaaaatgc agaataccta agggtcgcgc caciaaagcag
3781 tgaatttatt ggagcatgac cacggaggat agtatgagcc ctaaaaatcc agactctttc
3841 gatacccagg accaagccac agcaggtcct ccattcccac agccatgccc gcattagctc
3901 ttagaccac agactggttt tgcaacgttt acaccgacta gccaggaagt acttccacct
3961 cgggcacatt ttgggaagtt gcattccttt gtcttcaaac tgtgaagcat ttacagaaac
4021 gcatccagca agaatttgt ccctttgagc agaaatttat ctttcaaaga ggtatatttg
4081 aaaaaaaaaa aaaaagtata tgtgaggatt tttattgatt ggggatcttg gagtttttca
4141 ttgtcgctat tgatttttac ttcaatgggc tcttccaaca aggaagaagc ttgctggtag
4201 cacttgctac cctgagttca tccaggccca actgtgagca aggagcacia gccacaagtc
4261 ttccagagga tgcttgattc cagtggttct gcttcaaggc ttccactgca aaacactaaa
4321 gatccaagaa ggcttcatg gccccagcag gccggatcgg tactgtatca agtcatggca
4381 ggtacagtag gataagccac tctgtccctt cctgggcaaa gaagaaacgg aggggatgaa
4441 ttcttcctta gacttacttt tgtaaaaatg tccccacggg acttactccc cactgatgga
4501 ccagtggttt ccagtcatga gcgttagact gacttgtttg tcttccattc cattgttttg
4561 aaactcagta tgccgcccc tcttgctgt catgaaatca gcaagagagg atgacacatc
4621 aaataataac tcggattcca gccacattg gattcatcag catttggaac aatagcccac
4681 agctgagaat gtggaatacc taaggataac accgcttttg ttctcgcaaa aacgtatctc
4741 ctaatttgag gctcagatga aatgcatcag gtcctttggg gcatagatca gaagactaca
4801 aaaatgaagc tgctctgaaa tctcctttag ccatcaccac aaccccccaa aattagtttg
4861 tgttacttat ggaagatagt tttctccttt tacttactt caaaagcttt ttactcaaag
4921 agtatatgtt ccctccagg tctgtccctt cagctgcccc caaacccccct ccttacgctt tgtcacacaa
4981 aaagtgtctc tgccctgagt catctattca agcacttaca gctctggcca caacagggca
5041 ttttacaggt gcgaatgaca gtagcattat gagtagtgtg aattcaggtg gtaaatatga
5101 aactaggggt tgaaattgat aatgctttca caacatttgc agatgtttta gaaggaaaaa
5161 agttccttcc taaaataatt tctctacaat tggaagattg gaagattcag ctagttagga
5221 gcccatTTTT tctaattctg tgtgtgccct gtaacctgac tggttaacag cagtcccttg
5281 taaacagtgt tttaaactct cctagtcaat atccaccca tccaatttat caaggaagaa
5341 atggttcaga aaatatTTTc agcctacagt tatgttcagt cacacacaca tacaaaatgt
5401 tccttttgct tttaaagtaa tttttgactc ccagatcagt cagagccccct acagcattgt
5461 taagaaagta tttgattttt gtctcaatga aaataaaact atattcattt cc (SEQ ID

```

NO: 53)

FIGURE 30B

46/115

EGFR (NM_005228)

RPSGTAGAALLALLAALCPASRALEEKVCQGTSNKLTQLGTF
DHFLSLQRMFNNCEVVLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLE
LQIIRGNMYEYENSALAVLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVE
IQWRDIVSSDFLSNMSMDFQNLHLSGCKCDPSCPNGSCWGAGEENCQKLTKIICAQQ
SGRCRGKSPSDCCHNQCAAGCTGPRESDCLVCRKFRDEATCKDTCPPMLYNPTTYQ
DVNPEGKYSFGATCVKKCPRNYVVDHGSVCVRACGADSYEMEEDGVRKCKKCEGPCR
VCNGIGIGIEFKDLSINATNIKHFKNCTSIGDLHILPVAFRGDSFTHTPPLDPQEL
ILKTVKEITGFLLIQAWPENRTDLHAFENLEIIRGRTKQHGGQFSLAVVSLNITSLGL
SLKEISDGDVVISGKNLCYANTINWKKLFGTSGQKTKIISNRGENSCKATGQVCHA
CSPEGCWGPEPRDCVSCRNVSRGRECVDKCNLLEGEPRFVENSECTQCHPECLPQA
NITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVMGENNTLVWKYADAGHVCHLCHPNC
YGCTGPGLEGCPPTNGPKIPSIATGMVGALLLLLVVALGIGLFMRRRHIVRKRTLRL
QERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKI
VAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPFGCL
DYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRLVHRDLAARNVLVKTPQHVKITDF
LAKLLGAEEKEYHAEGGKVPIKWMALLESILHRIYTHQSDVWSYGVTVWELMTFGSKP
DGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIIEFSKMA
DPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADAYLIPQQGFFSSPSTS
TPLLSLSATSNNSTVACIDRNLQSCPIKEDSFLQRYSSDPTGALTEDSIDDTFLP
PEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLNVTQPTC
NSTFDSPAHWAQKGSQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRVAPQS
EFIGA (SEQ ID NO:54)

FIGURE 30C

47/115

EPHB2 (NM_004442)

```

1  gccccgggaa  gcgcagccat  ggctctgcgg  aggctggggg  ccgcgctgct  gctgctgccg
61  ctgctcgccg  ccgtggaaga  aacgctaata  gactccacta  cagcgactgc  tgagctgggc
121  tggatggtgc  atcctccatc  aggggtggga  gaggtgagtg  gctacgatga  gaacatgaac
181  acgatccgca  cgtaccaggt  gtgcaacgtg  tttgagtcaa  gccagaacaa  ctggctacgg
241  accaagttaa  tccggcgccg  tggcgcccac  cgcacccacg  tggagatgaa  gttttcgggtg
301  cgtgactgca  gcagcatccc  cagcgtgcct  ggctcctgca  aggagacctt  caacctctat
361  tactatgagg  ctgactttga  ctcgccacc  aagaccttcc  ccaactggat  ggagaatcca
421  tgggtgaagg  tggataccat  tgcagccgac  gagagcttct  ccaggtgga  cctgggtggc
481  cgcgtcatga  aaatcaacac  cgaggtgcgg  agcttcggac  ctgtgtcccg  cagcggcttc
541  tacctggcct  tccaggacta  tggcggtgcg  atgtccctca  tcgctgtgcg  tgccttctac
601  cgcaagtgc  cccgcatcat  ccagaatggc  gccatcttcc  aggaaacctt  gtcgggggct
661  gagagcatat  cgctggtggc  tgcccggggc  agctgcatcg  ccaatgcgga  agaggtggat
721  gtacccatca  agctctactg  taacggggac  ggcgagtggc  tggtgcccat  cgggcgctgc
781  atgtgcaaag  caggcttcga  ggccgttgag  aatggcaccg  tctgcccagg  ttgtccatct
841  gggactttca  aggccaacca  aggggatgag  gcctgtacc  actgtcccat  caacagccgg
901  accacttctg  aaggggccac  caactgtgtc  tgccgcaatg  gctactacag  agcagacctg
961  gaccccttgg  acatgccctg  cacaaccatc  cctcccgcc  ccaggtctgt  gatttccagt
1021  gtcaatgaga  cctccctcat  gctggagtgg  accctcccc  gcgactccgg  aggcgagag
1081  gacctcgtct  acaacatcat  ctgcaagagc  tgtggctcgg  gccgggtgc  ctgcaccgc
1141  tgcggggaca  atgtacagta  cgcaccacgc  cagctaggcc  tgaccgagcc  acgcatttac
1201  atcagtgacc  tgctggccca  caccagtagc  accttcgaga  tccaggtctg  gaacggcggt
1261  actgaccaga  gcccttcttc  gcctcagttc  gcctctgtga  acatcaccac  caaccaggca
1321  gctccatcgg  cagtgtccat  catgcatcag  gtgagccgca  ccgtggacag  cattaccttg
1381  tcgtggtccc  agccggacca  gcccaatggc  gtgatcctgg  actatgagct  gcagtactat
1441  gagaaggagc  tcagttagta  caacgccaca  gccataaaaa  gcccaccaa  cacggtcacc
1501  gtgcagggcc  tcaaagccgg  cgccatctat  gtcttcagg  tgccggcacg  caccgtggca
1561  ggctacgggc  gctacagcgg  caagatgtac  ttccagacca  tgacagaagc  cgagtaccag
1621  acaagcatcc  aggagaagtt  gccactcatc  atcggtcct  cggccgctgg  cctggtcttc
1681  ctcaattgctg  tggttgtcat  cgccatcgtg  tgtaacagaa  gacgggggtt  tgagcgtgct
1741  gactcggagt  acacggacaa  gctgcaacac  tacaccagt  gccacatgac  cccaggcatg
1801  aagatctaca  tcgatccttt  cacctacgag  gaccccaacg  aggcagtgcg  ggagtttgcc
1861  aaggaaattg  acatctcctg  tgtcaaaatt  gagcaggtga  tcggagcagg  ggagtttggc
1921  gaggtctgca  gtggccacct  gaagctgcca  ggcaagagag  agatctttgt  ggccatcaag
1981  acgctcaagt  cgggctacac  ggagaagcag  cgccgggact  tcctgagcga  agcctccatc
2041  atggggcag  tcgaccatcc  caacgtcatc  cacctggagg  gtgtcgtgac  caagagcaca
2101  cctgtgatga  tcatcaccga  gttcatggag  aatggctccc  tggactcctt  tctccggcaa
2161  aacgatgggc  agttcacagt  catccagctg  gtgggcatgc  ttcggggcat  cgcagctggc
2221  atgaagtacc  tggcagacat  gaactatgtt  caccgtgacc  tggctgccc  caacatcctc
2281  gtcaacagca  acctggtctg  caaggtgtcg  gactttgggc  tctcacgctt  tctagaggac
2341  gatacctcag  accccaccta  caccagtgc  ctgggcggaa  agatccccat  ccgtggaca
2401  gcccgggaag  ccatccagta  ccggaagttc  acctcggcca  gtgatgtgtg  gagtacggc
2461  attgtcatgt  gggaggtgat  gtccatggg  gagcgccct  actgggacat  gaccaaccag
2521  gatgtaata  atgccattga  gcaggactat  cggctgccac  cgcccatgga  ctgcccagc
2581  gccctgcacc  aactcatgct  ggactgttgg  cagaaggacc  gcaaccaccg  gccaaagttc
2641  ggccaaaattg  tcaacacgct  agacaagatg  atccgcaatc  ccaacagcct  caaagccatg
2701  gcgcccctct  cctctggcat  caacctgcg  ctgctggacc  gcacgatccc  cgactacacc
2761  agctttaaca  cgggtggacga  gtggctggag  gccatcaaga  tggggcagta  caaggagagc
2821  ttcgccaatg  ccggttcac  ctcccttgac  gtctgtctc  agatgatgat  ggaggacatt
2881  ctccgggttg  gggtcacttt  ggctggccac  cagaaaaaaa  tcctgaacag  tatccaggtg
2941  atgcggggcg  agatgaacca  gattcagct  gtggaggttt  gacattcacc  tgctcggct
3001  cacctcttcc  tccaagcccc  gcccctctg  cccacgtgc  cggccctcct  ggtgctctat

```

FIGURE 31A

48/115

```

3061 ccactgcagg gccagccact cggcaggagg ccacggggcca cgggaagaac caagcgggtgc
3121 cagccacgag acgtcaccaa gaaaacatgc aactcaaacg acggaaaaaa aaagggaatg
3181 ggaaaaaaga aaacagatcc tgggaggggg cgaggaaatac aagggaatatt ttttaaagag
3241 gattctcata aggaaagcaa tgactgttct tgcgggggat aaaaaagggc ttgggagatt
3301 catgcatgt gtccaatcgg agacaaaagc agtttctctc caactccctc tgggaagggtg
3361 acctggccag agccaagaaa cactttcaga aaaacaaatg tgaaggggag agacaggggc
3421 cgcccttggc tcctgtccct gctgctcctc taggcctcac tcaacaacca agcgcctgga
3481 ggacgggaca gatggacaga cagccaccct gagaaccctt ctgggaaaat ctattcctgc
3541 caccactggg caaacagaag aatttttctg tctttggaga gtatttttaga aactccaatg
3601 aaagacactg tttctcctgt tggtcacag ggctgaaagg ggcttttgtc ctctgggtc
3661 agggagaacg cggggacccc agaaagggtc gccttcctga ggatgggcaa ccccgaggtc
3721 tgcagctcca ggtacatata acgcgcacag cctggcagcc tggccctcct ggtgccact
3781 ccgccagcc cctgcctcga ggactgatac tgcagtact gccgtcagct ccgactgccg
3841 ctgagaaggg ttgatcctgc atctgggttt gtttacagca attcctggac tcgggggtat
3901 tttggtcaca ggggtgggttt gggttagggg gtttggttgt tgggttggtt tttgtttttt
3961 ggtttttttt aatgacaatg aagtgcact ttgacatttc ctaccttttg aggacttgat
4021 ccttctccag gaagaagggt ctttctgctt actgacttag gcaatacacc aagggcgaga
4081 ttttatatgc acatttctgg atttttttat acggttttca ttgacactct tccctcctcc
4141 cacctgccac caggcctcac caaagcccac tgccatgggg ccactctgggc cattcagaga
4201 ctggagttag atttgggtgt ggagggggag gcgccaagg ggaggagctt ccactccag
4261 gactgttgat gaaagggaca gattgaggag gaagtgggct ctgaggctgc agggctggaa
4321 gtccttgccc acttcccact ctctgcccc aatctatcta gtacttccca ggcaaatagg
4381 cccctttgag gctcctgagt gccctcagat ggtcaaaacc cagttttccc tctgggagcc
4441 taaaccaggc tgcacggag gccaggacc ggatcattca ctgtgatacc ctgccctcca
4501 gagggtgccg tcagagacac gggcaagcat gcctcttccc ttccctggag agaaagtgtg
4561 tgatttctct cccacctcct tccccccacc agaccttgc tgggcctaaa ggtcttggtc
4621 atggggagcg cctcagtcta gggatctggc cacagactcc ctctgtgaa ccaacacaga
4681 caccacagca gagcaatcag ttagtgaatt g (SEQ ID NO:55)

```

FIGURE 31B

EPHB2 (NM_004442)

```

ALRRLGAALLLLPILAAVEETLMDSTTATAELGWMVHPPSGWE
VSGYDENMNTIRTYQVCNVFESSQNNWLRTKFIIRRGAAHRIHVEMKFSVRDCSSIPS
PGSCKETFNLYYYEADFDSATKTFPNWMENPWVKVDTIAADESFSQVDLGGVRMKIN
EVRSFPGVSRSGFYLAQDYGGCMSLIAVRVFYRKCPRIIQNGAIFQETLSGAESTS
VAARGSCIANAEVDVPIKLYCNGDGEWLVPIGRCMCKAGFEAVENGTVCRGCPST
KANQGDEACTHCPINSRTTSEGATNCVCRNGYYRADLDPLDMPCTTIPSAPQAVISS
NETSLMLEWTPPRDSGGREDLVYNIICKSCGSGRGACTRCGDNVQYAPRQLGLTEPR
YISDLLAHTQYTFEIQAVNGVTDQSPFSPQFASVNITTNQAAPSAVSIMHQVSRVTD
ITLSWSQPDQPNGVILDYELQYYEKELSEYNATAIKSPTNTVTVQGLKAGAIYVFQV
ARTVAGYGRYSGKMYFQTMTEAEYQTSIQEKLPLIIGSSAAGLVFLIAVVVIAIVCN
RRGFERADSEYTDKLQHYTSGHMTPGMKIYIDPFTYEDPNEAVREFAKEIDISCVKI
QVIGAGEFGEVCSGHLKLPKREIFVAIKTLKSGYTEKQRRDFLSEASIMGQFDHPN
IHLEGVVTKSTPVMII TEFMENGSLDSFLRQNDGQFTVIQLVGMLRGIAAGMKYLAD
NYVHRDLAARNILVNSNLVCKVSDFGLSRFLEDDTSDPTYTSALGGKIPIRWTAPEA
QYRKFTSASDVWSYGI VMWEVMSYGERPYWDMTNQDVINAIEQDYRLPPPMDCPSAL
QLMLDCWQKDRNHRPKFGQIVNTLDKMIRNPNSLKAMAPLSSGINLPLLDRTIPDYT
FNTVDEWLEAIKMGQYKESFANAGFTSFDVVSQMMMEDILRVGVTLAGHQKKILNSI
VMRAQMNQIQSVEV (SEQ ID NO:56)

```

FIGURE 31C

49/115

CRIPTO CR-1 (NM_003212)

```

1 ggagaatccc cggaaaaggct gagtctccag ctcaagggtca aaacgtccaa ggccgaaagc
61 cctccagttt cccctggacg ccttgctcct gcttctgcta cgaccttctg gggaaaacga
121 atttctcatt ttcttcttaa attgccattt tcgcttttagg agatgaatgt tttcctttgg
181 ctgtttttggc aatgactctg aattaaagcg atgctaacgc ctcttttccc cctaattggt
241 aaaagctatg gactgcagga agatggcccg cttctcttac agtgtgattt ggatcatggc
301 catttctaaa gtctttgaac tgggattagt tgccgggctg ggccatcagg aatttgctcg
361 tccatctcgg ggataacctg ccttcagaga tgacagcatt tggcccagg aggagcctgc
421 aattcggcct cggctcttccc agcgtgtgcc gcccatgggg atacagcaca gtaaggagct
481 aaacagaacc tgctgcctga atgggggaac ctgcatgctg gggctctttt gtgcctgccc
541 tccctccttc tacggacgga actgtgagca cgatgtgcgc aaagagaact gtgggtctgt
601 gcccacatgac acctggctgc ccaagaagtg ttccctgtgt aaatgctggc acggtcagct
661 ccgctgcttt cctcaggcat ttctaccggg ctgtgatggc cttgtgatgg atgagcacct
721 cgtggcttcc aggactccag aactaccacc gtctgcacgt actaccactt ttatgctagt
781 tggcatctgc ctttctatac aaagctacta ttaatcgaca ttgacctatt tccagaaata
841 caatttttaga tatcatgcaa atttcatgac cagtaaaggc tgctgctaca atgtcctaac
901 tgaaagatga tcattttgtag ttgccttaaa ataatgaata caatttccaa aatggctctct
961 aacatttctc tacagaacta cttcttactt ctttgccctg cctctcccca aaaaactact
1021 tctttttttca aaagaaagtc agccatatct ccattgtgcc taagtccagt gtttcttttt
1081 tttttttttt ttgagacgga gtctcactct gtcacccagg ctggactgca atgacgcgat
1141 cttgggttcac tgcaacctcc gcatccgggg ttcaagccat tctcctgcct aagcctccca
1201 agtaactggg attacaggca tgtgtcacca tgcccagcta atttttttgt attttagtag
1261 agatgggggt ttcaccatat tggccagtct ggtctcgaac tctgacctg tgatccatcg
1321 atcagcctct cgagtgtctg gattacacac gtgagcaact gtgcaaggcc tgggtgtttct
1381 tgatacatgt aattctacca aggtcttctt aatatgttct tttaaatgat tgaattatat
1441 gttcagatta ttggagacta attctaattg ggaccttaga atacagtttt gagtagagtt
1501 gatcaaaatc aattaaaata gtctctttaa aaggaaagaa aacatcttta aggggaggaa
1561 ccagagtgtc gaaggaatgg aagtcctatc gcgtgtgtgc agggagactg ggtaggaaag
1621 aggaagcaaa tagaagagag aggttgaaaa acaaaatggg ttacttgatt ggtgattagg
1681 tgggtgtaga gaagcaagta aaaaggctaa atggaagggc aagtttccat catctataga
1741 aagctatata agacaagaac tccccttttt ttcccaaagg cattataaaa agaatgaagc
1801 ctcttagata aaaaaattat acctcaatgt cccaacaag attgcttaat aaattgtgtt
1861 tcttccaagc tattcaattc ttttaactgt tgtagaagac aaaatgttca caatatattt
1921 agttgtaaac caagtgatca aactacatat tgtaaagccc atttttaaaa tacattgtat
1981 atatgtgtat gcacagtaaa aatggaaact atattgacct aaaaaaaaaa aaa (SEQ ID
NO:57)

```

FIGURE 32A

CRIPTO CR-1 (NM_003212)

```

DCRKMARFSYSVIWIMAISKVFELGLVAGLGHQEFARPSRGYL
FRDDSIWPQEEPAIRPRSSQRVPPMGIQHSKELNRTCCLNGGTCMLGSFCACPPSFY
RNCEHDVRKENCGSVPHDTWLPKKCSLCKCWHGQLRCFPQAFLPGCDGLVMDEHLVA
RTPELPPSARTTTFMLVGICLSIQSY (SEQ ID NO:58)

```

FIGURE 32B

50/115

Eprin B1 (NM_004429)

```

1 gagtagacag cacagcggca gcgaggaggag tctatgcgag ctggacagca gtgggaggtt
61 tgtgaggctc gcactggccg cagaccctcg ggctcgatcg cccgggagcc aggactcggc
121 gacgcgaggc tgccgggcta cccggccgag gcttcggggg cgcaaactaa tgggactggc
181 tcgctcggca gcatctcccc gctcttctaa gtacactgag cagggcccg cgtgaagtag
241 aagctgtccg ggggcgcgta gcccgagtc ccagtgtggc ccggaggaac ggagcccgtg
301 ccagggcggc ccagtcggga gcccggggac cgagcttgtg ctgtggggaa acccccactt
361 cttccaaggg acagcgatcc cgggacggtc gaggcgtcgg ggcggtcacc gagacctctg
421 cgggaagacc ccgtcgggga gagggcgcgc agccccgaag cgtctcggga agtcgagcgg
481 aatcgggcgg gatcaccggy gggcgagag ccccgctcgc gcctcgtgcy gcagcggaga
541 gcccaggaga acgagccctc gggggccgaa gcccatgccc ggggtggggg cggctgcccc
601 gtgagtcctc ctggccggcc gggcgagaa gagcgacacc gaagccggcg ggaggggagc
661 acttcaaggc cggcggctgc ggaggatggg cgocctgagcg gctccgagcg cagcgcggca
721 gaggaaggcg aggcgagcct tggtagaggag gcgccaaggg atcccgaagt gcagtctgcc
781 cccgggaaga tggctcggcc tgggcagcgt tggctcggca agtggcttgt ggcgatggtc
841 gtgtgggcgc tgtgccggct cgccacaccg ctggccaaga acctggagcc cgtatcctgg
901 agtcacctca accccaagtt cctgagtggy aagggttgg tgatctatcc gaaaattgga
961 gacaagctgg acatcatctg ccccgagca gaagcagggc ggccctatga gtactacaag
1021 ctgtacctgg tgcggcctga gcaggcagct gcctgtagca cagtctctga cccaacgtg
1081 ttggctacct gcaataggcc agagcaggaa atacgcttta ccatcaagtt ccaggagttc
1141 agccccaact acatgggcct ggagttcaag aagcaccatg attactacat tacctcaaca
1201 tccaatggaa gcctggaggg gctggaaaac cgggagggcg gtgtgtgccg cacacgcacc
1261 atgaagatca tcatgaaggt tgggcaagat cccaatgctg tgacgcctga gcagctgact
1321 accagcaggg ccagcaagga ggcagacaac actgtcaaga tggccacaca ggccccctgg
1381 agtcggggct ccctgggtga ctctgatggc agcatgaga ctgtgaacca ggaagagaag
1441 agtggccag gtgcaagtgg gggcagcagc ggggacctg atggcttctt caactccaag
1501 gtggcattgt tcgcggtgtg cgggtgccgg tgcgtcatct tctgtctcat catcatcttc
1561 ctgacggtcc tactactgaa gctacgcaag cggcaccgca agcacacaca gcagcggcg
1621 gctgccctct cgctcagtac cctggccagt cccaaggggg gcagtggcac agcgggcacc
1681 gagcccgagc acatcatcat tcccttacgg actacagaga acaactactg cccccactat
1741 gagaaggtga gtggggacta cgggcaccct gtctacatcg tccaagagat gccgccccag
1801 agcccgccga acatctacta caaggtctga gtgcccggca cggcctcagg cccccgaggg
1861 acagtcggcc tggaccggac ctctcctttc gccccacac cccctccctt tggcagctgt
1921 gccaccttt gtatttagtt ttgtagtttc ttggctttta taatccctct tttccctg
1981 cccctgggct tcggaggggg gtgcttgtgc ccctaacccc catgctcttg tgccttcccc
2041 ctctggccag gcctctgggc tccgtggggg cggccctctt tgggaaggcag ggctggacac
2101 tgatggacag caggcaggga gacagtccc tggccctgcc cctccctcgc ccccttgcc
2161 accttcccag gactgcttgt ccgctatcat cactgttttt aatgcttttg tgttcatttt
2221 ttagctgtca actcattttc atctgttttt tgaagaaaaa tggaaaaaat taaaaggcag
2281 cccctcccca ggctttgtga gcctggccca agccagtaca agagggcctg gggcacgatg
2341 tggtcagcca ggaagcatag gatgccattt cttttataga ttccttggtg tttctgggtg
2401 ggtaaggggg aggccagggc tgttcacgcc catgagggaa gaggaaagt cactgggca
2461 aggtgtccca cctccctc ctgaccctcc tacgaggtt atcctggcaa tggggtagtc
2521 actgccaccc ttccacacac acacacacac acacacacac aaaaaaaat ccttccttg
2581 tgggattctt gggcatctcc tgcctccctc actctcacgg taattaatgt cttaattggc
2641 tgttgccctg ggaacaggag agctgctgca ggcagatgac ctcatggggg gtggaggag
2701 gtgaggtgcc caggtggcta tttgccctgc agagctggga gtttcacccc cccccccac
2761 cctgttctct ccttaccttt ggcatccttt ggccctgggt tctgggccc ttcagaggtg
2821 gagacctaa cgggtataag accaggtggc ctgctccttt ctgtcccagt cttgggctgg ggcctggaaa
2881 ggtaaccccc acccaaccca gctcctgctg ctggccagc ccgctgtgca ctttgacccc agttccttgc
2941 gaggaagagg ctgcctgggg ctggccagc gtgcgcttg gtcgacact caggcactcc cagcagcgc
3001 cagcacggct gctaacagac tgccacttga caccatccac ctccacactg cctcctggcc agtgcccac
3061 atggaaggag ctggccctca caccatccac ctccacactg cctcctggcc agtgcccac
3121 ccagtgcca ggtgggagag ggagcagaac agccagcccc ttccaggtgg cagtcggaag
3181 ggtttttgtt tttgtttctg ttgccatttg tgtaaatact agtctttttt gaaaaaaat
3241 aatgtaaaga tgttttgtat aaactctgaa ttattttctt gttgcttttt tcttagaaaa
3301 aaatgagaac taaaaaaaaa aaattaacca catggaaaaa aaaaaa (SEQ ID NO:59)

```

FIGURE 33A

51/115

Eprin B1 (NM_004429)

MARPGQRWLGKWLIVAMVVWALCRLATPLAKNLEPVSWSSLNPKF
LSGKGLVIYPKIGDKLDIICPRAEAGRPYEYYKLYLVRPEQAAACSTVLDPNVLVTCN
RPEQEIRFTIKFQEFSPNYMGLEFKKHHDYIITSTSNGLSLEGLNREGGVCRTMTKI
IMKVGQDPNAVTPQLTTSRPSKEADNTVKMATQAPGSRGSLGDSGKHETVNQEEKS
GPGASGGSSGDPDGGFFNSKVALFAAVGAGCVIFLLIIIFLTVLLLKLKRHRKHTQQR
AAALSLSTLASPKGGSGTAGTEPSDIIIPLRTTENNYCPHYEKVSGDYGHPVYIVQEM
PPQSPANIYYKV (SEQ ID NO:60)

FIGURE 33B

52/115

MMP-17/MT4-MMP (NM_016155)

```

1  ccggcggggg cgccgcggag agcggagggc gccgggctgc ggaacgcgaa ggggagggcg
61  cgggaccctg cacgccgccc gggggcccat gtgagcgcca tgcggcgccg cgcagcccgg
121  ggacccggcc cgccgcccc cggggccgga ctctcgcggt tgcgctgctt gccgctgccg
181  ctgctgctgc tgcctggcgt ggggaccgcg gggggctgct ccgcgcccgc acccgcgccg
241  cgcgcgagag acctcagcct gggagtggag tggctaagca ggttcggtta cctgcccccg
301  gctgacccca caacagggca gctgcagacg caagaggagc tgtctaaggc catcacagcc
361  atgcagcagt ttggtggcct ggaggccacc ggcatacctg acgaggccac cctggccctg
421  atgaaaaccc cacgctgctc cctgccagac ctccctgtcc tgaccaggc tgcaggaga
481  cgccaggctc cagccccac caagtggaac aagaggaaac tgcgtggag ggtccggacg
541  ttcccacggg actcaccact ggggcacgac acggtgcgtg cactcatgta ctacgccctc
601  aaggtctgga gcgacattgc gccctgaac ttccacgagg tggcgggcag caccgccgac
661  atccagatcg acttctccaa ggccgaccat aacgacggct accccttcga cggccccggc
721  ggcaccgtgg cccacgcctt cttcccggc caccaccaca ccgcgggga caccacttt
781  gacgatgacg aggcctggac cttccgctcc tcggatgccc acgggatgga cctgtttgca
841  gtggctgtcc acgagtttgg ccacgccatt ggggtaagcc atgtggccgc tgcacactcc
901  atcatgcggc cgtactacca gggcccgggt ggtgacccgc tgcgtacgg gctcccctac
961  gaggacaagg tgcgctctg gcagctgtac ggtgtgcggg agtctgtgtc tcccacggcg
1021  cagcccaggg agcctcccct gctgccggag ccccagaca accggtccag cggcccggcc
1081  aggaaggacg tgcccacag atgcagcact cactttgacg cggtgggcca gatccgggt
1141  gaagctttct tcttcaaagg caagtacttc tggcggtga cgcgggaccg gcacctggtg
1201  tccctgcagc cggcacagat gcaccgcttc tggcggggcc tgccgctgca cctggacagc
1261  gtggacgccc tgtacgagcg caccagcgac cacaagatcg tcttctttaa aggagacagg
1321  tactgggtgt tcaaggacaa taacgtagag gaaggatacc cgcgccccgt ctccgacttc
1381  agcctcccg cgtggcgcat cgacgtgccc ttctcctggg ccacaaatga caggacttat
1441  ttctttaaagg accagctgta ctggcgctac gatgaccaca cgaggacat ggaccccggc
1501  taccocggcc agagccccct gtggaggggt gtcocccagca cgctggacga cgccatgcgc
1561  tggctccgac gtgcctccta cttcttcggt gggcaggagt actggaaagt gctggatggc
1621  gagctggagg tggcacccgg gtaccacag tccacggccc gggactggct ggtgtgtgga
1681  gactcacagg ccgatggatc tgtggctgcg ggcgtggacg cggcagaggg gccccgcgcc
1741  cctccaggac aacatgacca gagccgctcg gaggacggtt acgaggtctg ctcatgcacc
1801  tctggggcat cctctcccc gggggcccca gggccactgg tggctgccac catgctgctg
1861  ctgctgccgc cactgtcacc aggcgcctcg tggacagcgg ccaggccctt gacgctatga
1921  cacacagcgc gagcccatga gaggacagag gcggtgggac agcctggcca cagagggcaa
1981  ggactgtgcc ggagtccctg ggggaggtgc tggcgcgga tgaggacggg ccaccctggc
2041  accggaaggc cagcagaggg cacggcccgc cagggtcggg caggctcagg tggcaaggac
2101  ggagctgtcc cctagtgagg gactgtgttg actgacgagc cgaggggtgg ccgctccaga
2161  aggggtgcca gtcaggccgc accgcgcgca gcctcctccg gccctggagg gagcatctcg
2221  ggctgggggg ccaccctctt ctgtgcgggc gccaccaacc ccaccacac tgctgcctgg
2281  tgctcccgcc ggcccacagg gcctccgtcc ccaggtcccc agtggggcag ccctcccac
2341  agacgagccc cccacatggt gccgcggcac gtccccctg tgacgcgttc cagaccaaca
2401  tgacctctcc ctgctttgta aaaaaaaaaa aaaaaaaa (SEQ ID NO:61)

```

FIGURE 34A

53/115

MMP-17/MT4-MMP (NM_016155)

MRRRAARGPGPPPPGGLSRLPLLPPLLLLLLALGTRGGCAAPA
PAPRAEDLSLGVWLSRFGYLPPADPTTGQLQTQEELSKATTAMQQFGGLEATGILDE
ATLALMKTPRCSLPDLFVLTQARRRRQAPAPTKWNKRNLNLSWRVRTFPRDSPLGHDTVR
ALMYYALKVWSDIAPLNFHEVAGSTADIQIDFSKADHNDGYPFDGPGGTVAHAFFPGH
HHTAGDTHFDDDEAWTFRSSDAHGMDFAVAVHEFGHAIGLSHVAAAHSIMRPYYQGP
VGDFLRYGLPYEDKVRVWQLYGVRESVSPTAQPEEPPLPEPPDNRSSAPPRKDVPHR
CSTHFDAVAQIRGEAFFFKGKYFWRLTRDRHLVSLQPAQMHRFWRGLPLHLDSVDVY
ERTSDHKIVFFKGDRYWVFKDNNVEEGYPRPVSDFSLPPGGIDAAFSWAHNDRTYFFK
DQLYWRYDDHTRHMDPGYPAQSPLWRGVPSTLDDAMRWS DGASYFFRGQEYWKVLDGE
LEVAPGYPQSTARDWLVC GDSQADGSVAAGVDAAEGPRAPPGQHDQSRSEDGYEVCSC
TSGASSPPGAPGPLVAATMLLLLLPPLSPGALWTAQAALTL (SEQ ID NO:62)

FIGURE 34B

54/115

MMP26 (NM_021801)

```
1 gacaaatgag ggtttggcat gcagctcgtc atcttaagag ttactatctt cttgccctgg
61 tgtttcgccg ttccagtgcc ccctgctgca gaccataaag gatgggactt tgttgagggc
121 tatttccatc aatttttctt gaccgagaag gagtcgccac tccttaccca ggagacacaa
181 acacagctcc tgcaacaatt ccatcggaat gggacagacc tacttgacat gcagatgcat
241 gctctgctac accagcccca ctgtggggtg cctgatgggt ccgacacctc catctcgcca
301 ggaagatgca agtggaataa gcacactcta acttacagga ttatcaatta cccacatgat
361 atgaagccat ccgcagtgaa agacagtata tataatgcag ttccatctg gagcaatgtg
421 acccctttga tattccagca agtgcagaat ggagatgcag acatcaaggt ttctttctgg
481 cagtgggccc atgaagatgg ttggcccttt gatgggccag gtggtatctt aggccatgcc
541 tttttaccaaa attctggaaa tcctggagtt gtccattttg acaagaatga aactgggtca
601 gcttcagaca ctggatataa tctgttcctg gttgcaactc atgagattgg gcattctttg
661 ggcctgcagc actctgggaa tcagagctcc ataatgtacc ccacttactg gtatcacgac
721 cctagaacct tccagctcag tgccgatgat atccaaagga tccagcattt gtatggagaa
781 aaatgttcat ctgacatacc ttaatgttag cacagaggac ttattcaacc tgtcctttca
841 gggagtttat tggaggatca aagaactgaa agcactagag cagccttggg gactgctagg
901 atgaagccct aaagaatgca acctagtcag gttagctgaa ccgacactca aaacgctact
961 gagtcacaat aaagattggt ttaaagagta aaaaaaaaaa aaaaaaaaaa (SEQ ID
```

NO:63)

FIGURE 35A

MMP26 (NM_021801)

```
MQLVILRVTFILPWCFVFPVPPAADHKGWDFVEGYFHQFFLTEK
SPLL TQETQTQLLQQFHRNGTDLLDMQMHALHQPCHGVPDGS DTSISPGRCKWNKH
LTYRIINYPHDMKPSAVKDSIYNAVSIWSNVTPLI FQQVQNGDADIKVSFWQWAHED
WPFDPGGGILGHAFLPNSGNPGVVHFDKNEHWSASDTGYNLFLVATHEIGHSLGLQH
GNQSSIMYPTYWYHDPRTFQLSADDIQR IQHLYGEKCSSDIP (SEQ ID NO:64)
```

FIGURE 35B

55/115

ADAM10 (NM_001110)

```

1 gaattcgagg atccgggtac catgggaggc ggcaggccta gcagcacggg aaccgtcccc
61 cgcgcgcgat cgcgcgcccc tgaagcgctt gggggacggg tatgggaggg aggtaggggc
121 gcggctccgc gtgccagttg ggtgcccgcg cgtcacgtgg tgaggaagga ggcggagggtc
181 tgagtttcga gggagggggg gagagaagag ggaacgagca agggaaggaa agcggggaaa
241 ggaggaagga aacgaacgag ggggaggagg gtccctgttt tggaggagct aggagcgttg
301 ccggccccctg aagtggagcg agaggagggt gcttcgccgt ttctcctgcc aggggagggtc
361 ccggcttccc gtggaggctc cggaccaagg cccttcagct tctccctccg gatcgatgtg
421 ctgctgttaa cccgtgagga ggcggcggcg gcggcagcgg cagcgggaaga tgggtgtgct
481 gagagtgtta attctgtctc tctcctgggc ggcggggatg ggaggtcagt atgggaatcc
541 tttaaataaa tatatcagac attatgaagg attatcttac aatgtggatt cattacacca
601 aaaacaccag cgtgccaaaa gagcagctctc acatgaagac caatttttac gtctagattt
661 ccatgcccat ggaagacatt tcaacctacg aatgaagagg gacacttccc ttttcagtga
721 tgaattttaa gtagaaacat caaataaagt acttgattat gatacctctc atatttacac
781 tggacatatt tatggtgaag aaggaagttt tagccatggg tctgttattg atggaagatt
841 tgaaggattc atccagactc gtgggtggcac attttatgtt gagccagcag agagatatat
901 taaagaccga actctgccat ttcactctgt catttatcat gaagatgata ttaactatcc
961 ccataaatac ggtcctcagg ggggctgtgc agatcattca gtatttgaaa gaataggaa
1021 ataccagatg actggtgtag aggaagtaac acagatacct caagaagaac atgctgctaa
1081 tgggtccagaa cttctgagga aaaaacgtac aacttcagct gaaaaaataa cttgtcagct
1141 ttatattcag actgatcatt tgttctttta atattacgga acacgagaag ctgtgattgc
1201 ccagatatcc agtcatgtta aagcgattga tacaatttac cagaccacag acttctccgg
1261 aatccgtaac atcagtttca tgggtgaaac cataagaatc aatacaactg ctgatgagaa
1321 ggaccctaca aatcctttcc gtttcccaaa tattggtgtg gagaagtttc tgggaattgaa
1381 ttctgagcag aatcatgatg actactgttt ggccatgtgc ttcacagacc gagatttga
1441 ttgtggcgta cttgggtctg cttgggttgc agcacctca ggaagctctg ggaagatag
1501 tgaaaaaagt aaactctatt cagatggtaa gaagaagtc ttaaactctg gaattattac
1561 tgttcagaac tatgggtctc atgtacctcc caaagtctct cacattactt ttgctcacga
1621 agttggacat aactttggat ccccatatga ttctggaaca gagtgcacac caggagaatc
1681 taagaatttg ggtcaaaaag aaaatggcaa ttacatcatg tatgcaagag caacatctgg
1741 ggacaaactt aacaacaata aattctcact ctgtagtatt agaaatataa gccaaagttct
1801 tgagaagaag agaaacaact gttttgttga atctggccaa cctatttgtg gaaatggaat
1861 ggtagaacaa ggtgaagaat gtgatttgtg ctatagtgc cagtgtaaag atgaatgctg
1921 ctctgatgca aatcaaccag agggaagaaa atgcaactg aaacctggga aacagtgcag
1981 tccaagtcaa ggtccttggt gtacagcaca gtgtgcattc aagtcaaagt ctgagaagtg
2041 tcgggatgat tcagactgtg caagggaagg aatatgtaat ggcttcacag ctctctgccc
2101 agcatctgac cctaaaccaa acttcacaga ctgtaatagg catacacaag tgtgcattaa
2161 tgggcaatgt gcaggttcta tctgtgagaa atatggctta gaggagtgtg cgtgtgccag
2221 ttctgatggc aaagatgata aagaattatg ccatgtatgc tgtatgaaga aaatggacc
2281 atcaacttgt gccagtacag ggtctgtgca gtggagtagg cacttcagtg gtcgaaccat
2341 caccctgcaa cctggatccc cttgcaacga ttttagagggt tactgtgatg ttttcatgcg
2401 gtgcagatta gtagatgctg atggctctct agctaggctt aaaaaagcaa tttttagtcc
2461 agagctctat gaaaacattg ctgaatggat tgtggctcat tgggtggcag tattacttat
2521 gggaattgct ctgatcatgc taatggctgg atttattaag atatgcagtg ttcatactcc
2581 aagtagtaat ccaaagttgc ctctcctaa accacttcca ggcactttaa agaggaggag
2641 acctccacag ccatttcagc aaccccagcg tcagcggccc cgagagagtt atcaaattggg
2701 acacatgaga cgctaactgc agcttttgcc ttgggtcttc ctagtgccta caatgggaaa
2761 acttcactcc aaagagaaac ctattaagtc atcatctcca aactaaacct tcacaagtaa
2821 cagttgaaga aaaaatggca agagatcata tcctcagacc aggtggaatt acttaaat
2881 taaagcctga aaattccaat ttgggggtgg gaggtggaaa aggaacccaa ttttcttatg
2941 aacagatatt ttttaactta tggcacaaag tcttagaata ttattatgtg ccccggtgtc
3001 cctgttcttc gttgtgtcat ttcttctact tgcaggcaaa cttggctctc aataaacttt
3061 taccacaat tgaaataaat atattttttt caactgcaa tcaaggctag gaggtcgcac
3121 caccctaaca ttggagacat cacttgccaa tgtacatacc ttgttatatg cagacatgta
3181 tttcttacgt acactgtact tctgtgtgca attgtaaaca gaaattgcaa tatggatgtt
3241 tctttgtatt ataaaatttt tccgctctta attaaaaatt actgtttaat tgacatactc
3301 aggataacag agaattggtg tattcagtgg tccaggattc tgtaatgctt tacacaggca
3361 gttttgaaat gaaaatcaat ttaccccatg gtaccoggat cctcgaattc (SEQ ID

```

NO: 65)

FIGURE 36A

56/115

ADAM10 (NM_001110)

VLLRVLILLLSWAAGMGGQYGNPLNKYIRHYEGLSYNVDSLHQ
HQRKRAVSHEDQFLRLDFHAHGRHFNLRMKRDTSLFSDEFKVETSNKVLDDYDTSHI
TGHIYGEEGSFSGHGSVIDGRFEGFIQTRGGTFYVEPAERYIKDRTLPPFHSVIYHEDD
NYPHKYGPQGGCADHSVFERMRKYQMTGVVEVTQIPQEEHAANGPELLRKKRTTSAE
NTCQLYIQTDHLFFKYYGTREAVIAQISSHVKAIDTIYQTTDFSGIRNISFMVKRIR
NTTADKDPNTNPFRRPNIGVEKFLELNSEQNHDDYCLAYVFTDRDFDDGVLGLAWVG
PSGSSGGICEKSKLYSDGKKKSLNTGIITVQNYGSHVPPKVSHITFAHEVGHNFSGP
DSGTECTPGESKNLGQKENGNYIMYARATSGDKLNNNKFSLCSIRNISQVLEKKRNN
FVESGQPICGNGMVEQGEECDGYSQCKDECCFDANQPEGRKCKLKPGKQCSPSQG
CCTAQCAFKSKSEKCRDDSDCAREGICNGFTALCPASDPKPNFTDCNRHTQVCINGQ
AGSICEKYGLEECTCASSDGKDDKELCHVCCMKMDPSTCASTGQSVQWSRHFSGRTI
LQPGSPCNDFRGYCDVFMRCRLVDADGPLARLKKAI FSPELYENIAEWIVAHWWAVL
MGIALIMLMAGFIKICSVHTPSSNPPLPPKPLPGTLKRRRPPQPIQQPQQRQRPRES
QMGHMRR (SEQ ID NO:66)

FIGURE 36B

57/115

ADAM1 (XM_132370)

```

1  cttgggtggg cagtgcaggc caactgcagt cagcaagtgt gggggcttaa gagttcttcc
61  agagcccaact tccattttct ttgttgcttt aactagagtc accagtctgt cttcattttt
121 atggtgagac cattgggaga actaacttag attttaggct ctaatatagt tctgtggtaa
181 aaataagatc atgtaacact tatgcttttag aaatttccat agagaaggat catgtcttaa
241 agccaaaatt tatttggttag acacaaggat acgggaaagt agaacatcta aatactgtgt
301 gtgtgtgcgt gtgcgtgtgc gtgtgtgtgt acaccagtga aaggaatcag gcagtctaag
361 agaactagct atccatccag catgaccact gtaagaatga ggaatgaggc aggacaacag
421 agaactctta attgttcaga gaaccagag aactttgtcc cctccccga aacctgcag
481 aatgttgagt ctgaaagtat gagctgggta acatgtcagg ggcccatgac ctgtggagga
541 ggaaagatga tgtgacaagc acagaaccgg ctgagccact gtagatgcag ggctcatctc
601 catgaatgtc aaaggaaact aagcaacact gaagctctc cacttgaaag aagccctgt
661 gctgcacata tccaccaagg ccaggagaaa gaaaggagag agacacagcc tgagaccgca
721 cagtttcttg ggaagctccc cagtaaggca cgggcacagg tctgggtgcc tgggtctggg
781 aaaagcagag agcactgccg ctgatggaca gagatcctcc atcatcagca gtttgttggg
841 gccatgtcag tggcagcagc ggggagaggg tttgcctcca gtctgtcttc cccacagatc
901 aggcgaatag ccttaaaaga agctaagcta acacctcaca tctgggcggc actgcactgg
961 aacttgggac tgagactagt gccatctgtc agagtaggga ttttgggtgt actgattttt
1021 cccccgagca cgttctgtga cattggatct gtatataaatt cttcctatga aactgtcatc
1081 cctgagagac tgccaggcaa gggggggaaa gaccctggag ggaaggtgtc ctacatgcta
1141 ttgatgcaag gccaaaagca gctgcttcac ctogaggtaa agggacacta ccctgagaat
1201 aacttcccag tctacagtta ccacaatggc atcctgaggc aagaaatgcc tctcctctcc
1261 caggactgcc actatgaagg ctacatggaa ggggtgccag gctcctttgt ttctgtcaac
1321 atctgttcag gcctcagggg ggtcttgatt aaagaggaaa catcctatgg cattgagccc
1381 atgctctctt ccaaaaactt tgaacatgtc ctctacacca tggagcatca gcctgtggtc
1441 tcctgcagtg tcaactccaa agacagccct ggggacacca gccatccacc aaggagcagg
1501 aagcccgatg acctactggg tctgactgac tgggtgtcac acaccaagta tgtggagatg
1561 tttgtgggtg tcaaccacca gcggttccag atgtggggca gtaacatcaa cgagacggtc
1621 caggcagtaa tggacatcat tgctctggcc aacagcttca ctagggggat aaacacagag
1681 gtgggtgctg tgggcctgga aatctggaca gagggggacc cgatagaggt cccagtggac
1741 ctgcagacca cactcaggaa tttcaacttc tggagacagg agaaactcgt gggccgggtc
1801 aggcacgatg tggcacactt gatcgtcggg catcgcccag gagagaacga gggccaggcg
1861 tttctccgtg gtgcctgttc gggtagtatt gggcgggccg tggaggcctt ccatcatgaa
1921 gatgtcctcc tgttcgcggc tctcatggcc cagagctcg ggcacaacct gggatatccag
1981 cagcaccacc cgacctgcac ctgtggctcc aagcacttct gcctcatggg tgagaagatc
2041 ggtaaggaca gtggcttcag caactgcagc tctgaccact tccctcggtt cctccatgac
2101 cacagagggg cgtgcctgct tgatgagcct gggcgccaga gccgcatgcg cagagctgcc
2161 aattgtggga atgggtgtgg ggaggacttg gaggagtgtg actgcggcag tgactgtgac
2221 agtcaccctg gctgttcgcc aacatgtacg ctttaaggagg gtgcgcagtg cagtgaggga
2281 ctctgctgct acaactgtac attcaagaag aaaggagct tatgccgtcc tgctgaggat
2341 gtgtgtgacc ttcccagata ttgtgacggc agtactcagg aatgccctgc aaacagctac
2401 atgcaggatg gcacacagtg tgataggatt tattactgct tggggggttg gtgtaagaac
2461 cctgataaac aatgttcaag gatctatggg tatcctgcaa gatctgcccc tgaggaatgt
2521 tacatttcag ttaatactaa ggcgaaccgg tttggaaact gtggccatcc cacctccgt
2581 aacttcagat atgaaacatg ttccgatgag gatgtatttt gtgggaaact ggtgtgtaca
2641 gatgttagat acctgccccaa agtcaaaccc ctacactcac tccctcaggt tccttatgga
2701 gaggactggg gttggagtat ggatgctat aacatcacag atgtcccgga tgacggagat
2761 gtacagagcg gcaccttctg tgccccaaac aaagtctgca tggagtatat ctgcactggg
2821 cgtgggggtg tccagtacaa ctgtgagcca caggaaatgt gtcacgggaa tggagtgtgc
2881 aacaatttca agcactgtca ctgcgatgct ggcttcgccc ctctgactg tagcagtcca
2941 ggaaatgggg ggagtgtgga cagtggctct gttggtaagc ccgctgatcg acacttgagt
3001 ctctcttttc tggctgaaga gagtccagat gataaaatgg aggatgaaga ggtaaacctg
3061 aaagtgatgg tgcttgtggg cctataatct ctgtctgttt tactgtgctg tctaagtctg
3121 atcgcttacc tctggctcga agtacaagaa ctgatatctc caccgagttc atcagagtct
3181 tctgtctcat catcctgggtc agactctgac tctcagtga gttttattta agatcctctc
3241 atggatcatt gctatcgatg tcttgtatct gcagggcaat tttgcctaag tggattttag
3301 ggcagtctgt tcagtgtaat gtgtgggtcta tatacttgtg ttgctcatct cagaacaac
3361 tgggaattata tctgaatga tgttaaggga tctaaatgtt ctaacttgcc ctgtcagctc
3421 ctgttcataa aatagaaggc attttaataa aatataaa (SEQ ID NO:67)

```

FIGURE 37A

58/115

ADAM1 (XM_132370)

MSVAAAGRGFASSLSSPQIRRIALKEAKLTPhiWAALHWNLGLR
LVPSVRVGILVLLIFLPSTFCDIGSVYNSSYETVIPERLPGKGGKDPGGKVSYMLLMQ
GQKQLLHLEVKGHY PENNFPVYSYHNGILRQEMPLLSQDCHYEGYMEGVPGSFVSVNI
CSGLRGVLIKEETSYGIEPMLSSKNFEHVLVTMEHQPVVSCSVTPKDSPGDTSHPPRS
RKPDDLVLTDWWSHTKYVEMFVVVNHQRFQMWGSNINETVQAVMDIIALANSFTRGI
NTEVVLVGLEIWTEDPIEVPVDLQTTLRNFNFWRQEKLVGRVRHDVAHLIVGHRPGE
NEGQAFLRGACSGEFAAAVEAFHHEDVLLFAALMAHELGHNLGIQHDHPTCTCGPKHF
CLMGEKIGKDSGFSNCSSDHFLRFLHDHRGACLLDEPGRQSRMRAANCGNGVVEDLE
ECDCGSDCDSHPCCSPTCTLKEGAQCSEGLCCYNCTFKKKGS LCRPAEDVCDLPEYCD
GSTQEC PANSYMQDGTQCDRIYYCLGGWCKNPDKQCSRIYGYPARSAPEECYISVNTK
ANRFGNCGHPTSANFRYETCSDDEDVFCGKLVCTDVRYLPKVKPLHSL LQVPYGEDWCW
SMDAYNITDVPDDGDVQSGTFCAPNKVCM EYICTGRGVLQYNCEPQEMCHGNGVCNNF
KHCHCDAGFAPPDCSSPGNGGSVD SGPVGKPADRHL SLSFLAEESPDDKMEDEEVNLK
VMVLVVP IFLVVL LCCMLLIAYLWSEVQEVVSPSSSESSSSSSWSDSDSQ (SEQ ID NO:68)

FIGURE 37B

59/115

TIM1 (NM_003254)

```
1  agggggcctta  gcgtgccgca  tcgccgagat  ccagcgccca  gagagacacc  agagaaccca
61  ccatggcccc  ctttgagccc  ctggcttctg  gcatcctggt  gttgctgtgg  ctgatagccc
121 ccagcagggc  ctgcacctgt  gtcccacccc  acccacagac  ggcttctgc  aattccgacc
181 tcgtcatcag  ggccaagttc  gtggggacac  cagaagtcaa  ccagaccacc  ttataccagc
241 gttatgagat  caagatgacc  aagatgtata  aagggttcca  agccttaggg  gatgccgctg
301 acatccgggt  cgtctacacc  cccgccatgg  agagtgtctg  cggatacttc  cacagggtcc
361 acaaccgcag  cgaggagttt  ctcattgctg  gaaaactgca  ggatggactc  ttgcacatca
421 ctacctgcag  tttcgtggct  ccctggaaca  gcctgagctt  agctcagcgc  cggggcttca
481 ccaagaccta  cactgttggc  tgtgaggaat  gcacagtgtt  tccctgttta  tccatcccc
541 gcaaactgca  gagtggcact  cattgcttgt  ggacggacca  gtcctccaa  ggctctgaaa
601 agggcttcca  gtcccgtcac  cttgcctgcc  tgccctcgga  gccagggctg  tgcacctggc
661 agtccctgcg  gtcccagata  gcctgaatcc  tgcccgaggt  ggaactgaag  cctgcacagt
721 gtccaccctg  ttcccactcc  catctttctt  ccggacaatg  aaataaagag  ttaccacca
781 gc (SEQ ID NO:69)
```

FIGURE 38A

TIM1 (NM_003254)

```
APFEPLASGILLLLWLIAPSRACCTCVPPHPQTAFCNSDLVIRA
FVGTPEVNQTTLYQRYEIKMTKMYKGFQALGDAADIRFVYTPAMESVCGYFHRSHNR
EEFLIAGKLQDGLLHITTCFVAPWNSLSLAQRRGFTKTYTVGCEECTVFPCLSI PC
LQSGTHCLWTDQLLQSEKGFQSRHLACLPREPGLCTWQSLRSQIA (SEQ ID NO:70)
```

FIGURE 38B

60/115

MUC1 (XM_053256)

```

1  cgctccacct ctcaagcagc cagcgccctgc ctgaatctgt tctgccccct cccaccccat
61  ttcaccacca ccatgacacc gggcaccacag tctcctttct tctgctgct gctcctcaca
121 gtgcttacag ttgttacagg ttctgggtcat gcaagctcta cccaggttg agaaaaggag
181 acttcggcta cccagagaag ttcagtgcgc agctctactg agaagaatgc tgtgagtatg
241 accagcagcg tactctccag ccacagcccc ggttcaggct cctccaccac tcagggacag
301 gatgtcactc tggccccggc cacggaacca gcttcagggt cagctgccac ctggggacag
361 gatgtcacct cggctcccagt caccaggcca gccctgggct ccaccacccc gccagcccac
421 gatgtcacct cagccccgga caacaagcgg gcccggggct ccaccgcccc cccagcccac
481 ggtgtcacct cggccccgga caccaggccg gcccggggct ccaccgcccc cccagcccac
541 ggtgtcacct cggccccgga caacaggccc gccttgggct ccaccgcccc tccagtccac
601 aatgtcacct cggcctcagg ctctgcatca ggctcagctt ctactctggt gcacaacggc
661 acctctgcca gggctaccac aaccccagcc agccaagagca ctccattctc aattcccage
721 caccactctg atactcctac cacccttgcc agccatagca ccaagactga tgccagtagc
781 actcaccata gcacgggtacc tctcttcacc tctccaatc acagcacttc tccccagttg
841 tctactgggg tctctttctt tttcctgtct tttcacattt caaacctcca gtttaattcc
901 tctctggaag atcccagcac cgactactac caagagctgc agagagacat ttctgaaatg
961 tttttgcaga tttataaaca aggggggttt ctggggcctct ccaatattaa gttcaggcca
1021 ggatctgtgg tggtaacaatt gactctggcc ttccgagaag gtaccatcaa tgtccacgac
1081 gtggagacac agttcaatca gtataaaacg gaagcagcct ctcgatataa cctgacgac
1141 tcagacgtca gcgtgagtga tgtgccattt cctttctctg cccagtctgg ggctgggggtg
1201 ccaggctggg gcacgcgct gctgggtgct gtctgtgttc tggttgcgct ggccattgtc
1261 tatctcattg ccttggctgt ctgtcagtgc cgcgaaaga actacgggca gctggacac
1321 tttccagccc gggataccta ccacctatg agcaggtacc ccacctacca caccatggg
1381 cgctatgtgc cccctagcag taccgatcgt agcccctatg agaaggtttc tgcaggtaat
1441 ggtggcagca gcctctctta cacaaaccca gcagtggcag ccacttctgc caactttag
1501 gggcacgtcg cccgtgagc tgagtggcca gccagtgcca ttccactcca ctcaggttct
1561 tcagggccag agcccctgca ccctgtttgg gctgggtgagc tgggagttca ggtgggctgc
1621 tcacagcctc cttcagaggc cccaccaatt tctcggacac ttctcagtgt gtggaagctc
1681 atgtgggccc ctgagggtc atgcctggga agtgtgtgg tgggggctcc caggaggact
1741 ggcccagaga gccctgagat agcggggatc ctgaactgga ctgaataaaa cgtgggtctcc
1801 cactg (SEQ ID NO:71)

```

FIGURE 39A

MUC1 (XM_053256)

```

MTPGTQSPFFLLLLLLTVLTVVTGSGHASSTPGGEKETSATQRSS
VPSSTEKNAVSMTSSVLSSHSPGSGSSTTQGDVTLAPATEPASGSAATWGQDVTSPV
VTRPALGSTTPPAHDVTSAPDNKRARGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS
APDNRPALGSTAPPVHNVTASGSASGSASTLVHNGTSARATTPASKSTPFSIPSHH
SDTPTTLASHSTKTDASSTHHSTVPPLTSSNHSTSPQLSTGVSFFFLSFHISNLQFNS
SLEDPSTDYQELQORDISEMFLQIYKQGGLGLSNIKFRPGSVVVQLTLAFREGTINV
HDVETQFNQYKTEAASRYNLTISDVSVDVPPFSAQSGAGVPGWGIALLVLCVLVA
LAIVYLIALAVCQCRKKNYGQLDIFPARDTYHPMSEYPTYHTHGRYVPPSSSTRSPYE
KVSAGNGGSSLSYTNPAVAATSANL (SEQ ID NO:72)

```

FIGURE 39B

61/115

CEA (NM_004363)

```

1  ctcagggcag agggaggaag gacagcagac cagacagtca cagcagcctt gacaaaacgt
61  tcoctggaact caagctcttc tccacagagg aggacagagc agacagcaga gaccatggag
121 tctccctcgg cccctcccca cagatggtgc atccctcggc agaggctcct gctcacagcc
181 tcaattctaa ccttctggaa cccgcccacc actgccaagc tcactattga atccacgcg
241 ttcaatgtcg cagaggggaa ggagggtgctt ctacttgtcc acaatctgcc ccagcatctt
301 tttgggtaca gctggtacaa aggtgaaaga gtggatggca accgtcaaat tataggatat
361 gtaataggaa ctcaacaagc taccacaggg cccgcataca gtggctcgaga gataatatac
421 cccaatgcat ccctgctgat ccagaacatc atccagaatg acacaggatt ctacacccta
481 cacgtcataa agtcagatct tgtgaatgaa gaagcaactg gccagttccg ggtatacccg
541 gagctgcca agccctccat ctccagcaac aactccaaac ccgtggagga caaggatgct
601 gtggccttca cctgtgaacc tgagactcag gacgcaacct acctgtggtg ggtaaacaat
661 cagagcctcc cggtcagtcc caggctgcag ctgtccaatg gcaacaggac cctcactcta
721 ttcaatgtca caagaaatga cacagcaagc taaaaatgtg aaaccagaa cccagtgagt
781 gccaggcgca gtgattcagt catcctgaat gtccctctatg gcccgatgc ccccaccatt
841 tcccctctaa acacatctta cagatcaggg gaaaatctga acctctcctg ccacgcagcc
901 tctaaccacac ctgcacagta ctcttgggtt gtcaatggga ctttcagca atccacccaa
961 gagctcttta tcccacacat cactgtgaat aatagtggat cctatacgtg ccaagcccat
1021 aactcagaca ctggcctcaa taggaccaca gtcacgacga tcacagtcta tgcagagcca
1081 cccaaaccct tcatcaccag caacaactcc aaccccgtgg aggatgagga tgctgtagcc
1141 ttaacctgtg aacctgagat tcagaacaca acctacctgt ggtgggtaaa taatcagagc
1201 ctcccgggtca gtcccaggct gcagctgtcc aatgacaaca ggacctcac tctactcagt
1261 gtcacaagga atgatgtagg accctatgag tgtggaatcc agaacgaatt aagtgttgac
1321 cacagcgacc cagtcacctc gaatgtctc tatggcccag acgacccac catttcccc
1381 tcatacacct attaccgtcc aggggtgaac ctacgcctct cctgccatgc agcctctaac
1441 ccacctgcac agtattcttg gctgattgat gggaacatcc agcaacacac acaagagctc
1501 tttatctcca acatcactga gaagaacagc ggactctata cctgccaggc caataactca
1561 gccagtggtc acagcaggac tacagtcaag acaatcacag tctctgcgga gctgcccagg
1621 cctccatct ccagcaacaa ctccaaaccc gtggaggaca aggatgctgt ggccttcacc
1681 tgtgaacctg aggtcagaa cacaacctac ctgtggtggg taaatggtca gagcctccca
1741 gtcagtccca ggctgcagct gtccaatggc aacaggacc cactctatt caatgtcaca
1801 agaaatgacg caagagccta tgtatgtgga atccagaact cagtgagtgc aaaccgcagt
1861 gaccagtcga cctggatgt cctctatggg ccggacaccc ccatcatttc cccccagac
1921 tcgtcttacc tttcgggagc gaacctcaac ctctcctgcc actcggcctc taacctatcc
1981 ccgcagtatt cttggcgtat caatgggata ccgcagcaac acacacaagt tctctttatc
2041 gccaaaatca cgccaaataa taacgggacc tatgcctgtt ttgtctctaa cttggctact
2101 ggccgcaata attccatagt caagagcatc acagtctctg catctggaac ttctcctggt
2161 ctctcagctg gggccactgt cttcatttca ggaagactga cagttgtttt gcttcttctc
2221 tagcagccct ggtgtagttt gcaacagcta cagtctaaaa ttgtctcttt accaaggata tttacagaaa
2281 taaagcattt agactctgac cagagatcga gaccatccta gccaacatcg tgaaacccca tctctactaa
2341 agactctgac caagatcgga atgagctggg cttggtggcg cgcacctgta gtcccagtta ctccggaggc
2401 aaatacaaaa tgagctggg gaatcgcttg aaccggggag gtggagattg cagtgagccc agatcgacc
2461 tgaggcagga agtctggcaa cagagcaaga ctccatctca aaaagaaaag aaaagaagac
2521 actgcactcc actcttgaat acaagtttct gataccactg cactgtctga gaatttccaa
2581 tctgacctgt aactaactga cagcttcatg aaactgtcca ccaagatcaa gcagagaaaa
2641 aactttaatg aactaactga catgggacta aatgaactaa tgaggattgc tgattcttta aatgtcttgt
2701 taattaattt catgggacta ttttttcttt taagctatcc actcttacag caatttgata
2761 tcccagatt tcaggaaact ttttttcttt atttacattt tctccctatg tggctcgctc
2821 aaatatactt ttgtgaacaa aaattgagac atttatattt aatatagtta ttgcacaagt
2881 agacttggga aactattcat gaatatattt attgtatggt aatatagtta ttgcacaagt
2941 tcaataaaaa tctgctcttt gtataacaga aaaa (SEQ ID NO:73)

```

62/115

CEA (NM_004363)

MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFN
VAEGKEVLLL VHNLPQH LFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREIIY
PNASLLIQNI IQNDTG FYTLHVIKSDLVNEEATGQFRVYPELPKPSISSNNSKPVEDK
DAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQ
NPVSARRSDSVILNVLYGPDAPTISPLNTSYRSGENLNLSCHAASNPPAQYSWFVNGT
FQQSTQELFIPNITVNNSGSYTCQAHNSDTGLNRTTVTTITVYAEPPKPFITSNNSNP
VEDEDAVALTCEPEIQNTTYLWWVNNQSLPVSPRLQLSNDNRTLTLLSVTRNDVGPYE
CGIQNELSVDHSDPVILNVLYGPDPTISPSYTYRPGVNL SLSCHAASNPPAQYSWL
IDGNIQQHTQELFISNITEKNSGLYTCQANNSASGHSRTTVKTITVSAELPKPSISSN
NSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQSLPVSPRLQLSNGNRTLTLFNVTRNDA
RAYVCGIQNSVSANRSDPVTLDVLYGPDTPII SPDSYLSGANLNLSCHSASNPS PQ
YSWRINGIPQQHTQVLFIAKITPNNNGTYACFVSNLATGRNNSIVKSITVSASGTSPG
LSAGATVGIMIGVLVGVALI (SEQ ID NO:74)

FIGURE 40B

63/115

NCA (NM_002483)

```

1  ctctcttaca aagaggtgga cagagaagac agcagagacc atgggacccc cctcagcccc
61  tccctgcaga ttgcatgtcc cctggaagga ggtcctgctc acagcctcac ttctaacctt
121 ctggaaccca cccaccactg ccaagctcac tattgaatcc acgccattca atgtcgcaga
181 ggggaaggag gttcttctac tcgcccacaa cctgccccag aatcgtattg gttacagctg
241 gtacaaaggc gaaagagtgg atggcaacag tctaattgta ggatatgtaa taggaactca
301 acaagctacc ccagggcccg catacagtgg tcgagagaca atatacccca atgcatccct
361 gctgatccag aacgtcaccc agaattctat accctacaag tcataaagtc
421 agatcttgtg aatgaagaag caaccggaca gttccatgta taccgggagc tgcccaagcc
481 ctccatctcc agcaacaact ccaaccccggt ggaggacaag gatgctgtgg ccttcacctg
541 tgaacctgag gttcagaaca caacctacct gtggtgggta aatggtcaga gcctcccggt
601 cagtcccagg ctgcagctgt ccaatggcaa catgacctc actctactca gcgtcaaaag
661 gaacgatgca ggatcctatg aatgtgaaat acagaaccca gcgagtgcc accgcagtga
721 ccagtcacc ctgaatgtcc tctatggccc agatgtcccc accatttccc cctcaaaggc
781 caattaccgt ccaggggaaa atctgaacct ctcccgccac gcagcctcta acccacctgc
841 acagtactct tggtttatca atgggacggt ccagcaatcc acacaagagc tctttatccc
901 caacatcact gtgaataata gcggtatccta tatgtgcaa gcccataact cagccactgg
961 cctcaatagg accacagtca cgatgatcac agtctctgga agtgctcctg tctctcagc
1021 tgtggccacc gtcggcatca cgattggagt gctggccagg gtggctctga tatagcagcc
1081 ctgggtgtatt ttcgatattt caggaagact ggcagattgg accagaccct gaattcttct
1141 agctcctcca atcccatttt atcccatgga accactaaaa acaaggctctg ctctgctcct
1201 gaagccctat atgctggaga tggacaactc aatgaaaatt taaagggaaa accctcaggc
1261 ctgaggtgtg tgccactcag agacttcacc taactagaga cagtcaaact gcaaaccatg
1321 gtgagaaatt gacgacttca cactatggac agcttttccc aagatgtcaa aacaagactc
1381 ctcatcatga taaggctctt accccctttt aatttgtcct tgcttatgcc tgccctttc
1441 gcttggcagg atgatgctgt cattagtatt tcacaagaag tagcttcaga gggtaactta
1501 acagagtgtc agatctatct tgtcaatccc aacgttttac ataaaataag agatccttta
1561 gtgcacccag tgactgacat tagcagcatc tttaacacag ccgtgtgttc aaatgtacag
1621 tggtcctttt cagagttgga cttctagact cacctgttct cactccctgt ttttaattcaa
1681 ccagcccatg caatgccaaa taatagaatt gctccctacc agctgaacag ggaggagtct
1741 gtgcagtttc tgacacttgt tgttgaaatc ggctaaatac aatgggtatc gctgagacta
1801 agttgtagaa attaacaaat gtgctgcttg gttaaaatgg ctacactcat ctgactcatt
1861 ctttatttcta ttttagttgg tttgtatctt gcctaagggt cgtagtccaa ctcttggtat
1921 taccctccta atagtcatac tagtagtcat actccctggg gtagtgtatt ctctaaaagc
1981 tttaaatgtc tgcatgcagc cagccatcaa atagtgaatg gtctctcttt ggctggaatt
2041 acaaaaactca gagaaatgtg tcatcaggag aacatcataa cccatgaagg ataaaagccc
2101 caaatggtgg taactgataa tagcactaat gctttaagat ttggtcacac tctcacctag
2161 gtgagcgcac tgagccagtg gtgctaaatg ctacatactc caactgaaat gttaaggaag
2221 aagatagatc caaaaaaaaa aaaaaaaaaa (SEQ ID NO:75)

```

FIGURE 41

64/115

NCA (NM_002483)

MGPPSAPPCRLHVPWKEVLLTASLLTFWNPPTTAKLTIESTPFN
VAEGKEVLLLAHNLPQNRIGYSWYKGERVDGNSLIVGYVIGTQQATPGPAYSGRETIY
PNASLLIQNVTQNDTGFTLQVIKSDLVNEEATGQFHVYPELPKPSISSNNSNPVEDK
DAVAFTCEPEVQNTTYLWWVNGQSLPVSPRLQLSNGNMTLTLLSVKRNDAGSYECEIQ
NPASANRSDPVTNLNVLYGPDVPTISPSKANYRPGENLNLSCHAASNPPAQYSWFINGT
FQOSTQELFIPNITVNNSGSYMCQAHNSATGLNRITTVTMITVSGSAPVLSAVATVGIT
IGVLARVALI (SEQ ID NO: 76)

FIGURE 41B

65/115

Follistatin (NM_006350)

```

1  gctcctcgcc ccgcgcctgc ccccaggatg gtccgcgcga ggcaccagcc ggggtgggctt
61  tgcctcctgc tgctgctgct ctgccagttc atggagggacc gcagtgccca ggctgggaac
121  tgctggctcc gtcaagcgaa gaacggccgc tgccagggtcc tgtacaagac cgaactgagc
181  aaggaggagt gctgcagcac cggccggctg agcacctcgt ggaccgagga ggacgtgaat
241  gacaacacac tcttcaagtg gatgattttc aacggggggcg cccccaactg catcccctgt
301  aaagaaacgt gtgagaacgt ggactgtgga cctgggaaaa aatgccgaat gaacaagaag
361  aacaaacccc gctgcgtctg cgccccggat tgttccaaca tcacctggaa ggggtccagtc
421  tgccggctgg atgggaaaaac ctaccgcaat gaatgtgcac tcctaaaggc aagatgtaaa
481  gagcagccag aactggaagt ccagtaccaa ggcagatgta aaaagacttg tcgggatggt
541  ttctgtccag gcagctccac atgtgtggtg gaccagacca ataatgccta ctgtgtgacc
601  tgtaatcgga tttgcccaga gcctgcttcc tctgagcaat atctctgtgg gaatgatgga
661  gtcacctact ccagtgcctg ccacctgaga aaggctacct gcctgctggg cagatctatt
721  ggattagcct atgagggaaa gtgtatcaaa gcaaagtcct gtgaagatat ccagtgcact
781  ggtgggaaaa aatgtttatg ggatttcaag gttgggagag gccggtgttc cctctgtgat
841  gagctgtgcc ctgacagtaa gtcggatgag cctgtctgtg ccagtgacaa tgccacttat
901  gccagcgagt gtgccatgaa ggaagctgcc tgctcctcag gtgtgctact ggaagtaaag
961  cactccggat cttgcaactg aatctgcccg taaaacctga gccattgatt cttcagaact
1021  ttctgcagtt tttgacttca tagattatgc tttaaaaaat tttttttaac ttattgcata
1081  acagcagatg ccaaaaaaaa aaaaagcatc tcaactgcaag tcacataaaa atgcaacgct
1141  gtaatattggc tgtatcagag ggctttgaaa acatacactg agctgcttct gcgctgttgt
1201  tgtccgtatt taaacaacag ctcccctgta ttcccccatc tagccatttc ggaagacacc
1261  gaggaagagg aggaagatga agaccaggac tacagctttc ctatatcttc tattctagag
1321  tggtaaactc tctataagtg ttcagtgttc acatagcctt tgtgcaaaaa aaaaaaaaaa
1381  aaaaaa (SEQ ID NO:77)

```

FIGURE 42A

66/115

Follistatin (NM_006350)

MVRARHQPGGLLLLLLLCQFMEDRSAQAGNCWLRQAKNGRCQV
LYKTELSKEECCSTGRLSTSWTEEDVNDNTLFWMI FNGGAPNCIPCKETCENVDCGP
GKKCRMNKKNKPRCVCAPDCSNITWKGPVCGLDGKTYRNECALLKARCKEQPELEVQY
QGRCKKTCRDVFCPGSSSTCVVDQTNNAYCVTCNRICPEPASSEQYLCGNDGVITYSSAC
HLRKATC LLGRSIGLAYEGKCIKAKSCEDIQCTGGKKCLWDFKVGRGRCSLCDELCPD
SKSDEPVCASDNATYASECAMKEAACSSGVLLLEVKHSGSCN (SEQ ID NO: 78)

FIGURE 42B

67/115

Claudin 1 (NM_021101)

```

1 gagcaaccgc agcttctagt atccagactc cagcgccgccc cggggcgccg accccaacccc
61 cgaccagag cttctccagc ggcggcgccg cgagcagggc tccccgcctt aacttcctcc
121 gcggggccca gccaccttcg ggagtcgggg ttgcccacct gcaaactctc cgccttctgc
181 acctgccacc cctgagccag cgcggggcgcc cgagcgagtc atggccaacg cggggctgca
241 gctgttgggc ttcattctcg ccttcctggg atggatcggc gccatcgta gcactgcctt
301 gccccagtgg aggatttact cctatgccgg cgacaacatc gtgaccggcc aggccatgta
361 cgaggggctg tggatgtcct gcgtgtcgca gagcaccggg cagatccagt gcaaagctt
421 tgactccttg ctgaatctga gcagcacatt gcaagcaacc cgtgccttga tgggtggttg
481 catcctcctg ggagtgatag caatctttgt ggccaccgtt ggcatgaagt gtatgaagtg
541 cttggaagac gatgaggtgc agaagatgag gatggctgtc attgggggtg cgatatttct
601 tcttgccagg ctggctattt tagttgccac agcatgggat ggcaatagaa tcgttcaaga
661 attctatgac cctatgaccc cagtcaatgc caggtaacgaa tttggtcagg ctctcttcac
721 tggctgggct gctgcttctc tctgccttct gggaggtgcc ctactttgct gttcctgtcc
781 ccgaaaaaca acctcttacc caacaccaag gccctatcca aaacctgcac cttccagcgg
841 gaaagactac gtgtgacaca gaggcaaaag gagaaaatca tgttgaaaca aaccgaaaat
901 ggacattgag atactatcat taacattagg accttagaat tttgggtatt gtaactgaa
961 gtatggtatt acaaaacaaa caaacaacaa aaaaacccat gtgttaaaat actcagtgc
1021 aaacatggct taatcttatt ttatcttctt tctcaatat agggagggaag atttttccat
1081 ttgtattact gcttcccat gagtaatcat actcaattgg ggggaaggggt gctccttaaa
1141 tatatataga tatgtatata tacatgtttt tctattaaaa atagacagta aaatactatt
1201 ctcatatagt tgatactagc atacttaaaa tatctctaaa ataggtaaat gtatttaatt
1261 ccatattgat gaagatgttt attggtatat tttcttttct gtctatata acatatgtaa
1321 cagtcaaata tcatttactc ttcttcatta gctttgggtg cctttgccac aagacctagc
1381 ctaattttacc aaggatgaat tctttcaatt cttcatgcgt gcccttttca tactattatt
1441 ttatttttta ccataatctt atagcacttg catcgttatt aagcccttat ttgttttg
1501 tttcattggg ctctatctcc tgaatctaac acatttcata gectacattt tagtttctaa
1561 agccaagaag aatttattac aaatcagaac tttggaggca aatctttctg catgaccaa
1621 gtgataaatt cctgttgacc ttcccacaca atccctgtac tctgacccat agcactcttg
1681 tttgctttga aaatatttgt ccaattgagt agctgcatgc tgttccccc ggtgttgtaa
1741 cacaacttta ttgattgaat tttaagcta cttattcata gttttatata cccctaaact
1801 acctttttgt tccccattcc ttaattgtat tgttttccca agtgtaatta tcatgcgttt
1861 tatactctcc taataagggtg tggctgtgtt gtctgaacaa agtgctagac tttctggagt
1921 gataatctgg tgacaaatat tctctctgta gctgtaagca agtcacttaa tcttctacc
1981 tcttttttct atctgccaaa ttgagataat gatacttaac cagttagaag aggtagtgtg
2041 aatattaatt agtttatatt actctcattc tttgaacatg aactatgcct atgtagtgtc
2101 tttatttgct cagctggctg agacactgaa gaagtcactg aacaaaacct acacacgtac
2161 cttcatgtga ttcactgcct tctctctctc accagtctat tccactgaa caaacctac
2221 acacatacct tcatgtggtt cagtgccttc ctctctctac cagtctattt ccactgaaca
2281 aaacctacgc acataccttc atgtggctca gtgccttctc ctctctacca gtctatttcc
2341 attctttcag ctgtgtctga catglttggt ctctgttcca ttttaacaac tgctcttact
2401 tttccagtc gtacagaatg ctatttcact tgagcaagat gatgtaatgg aaagggtgtt
2461 ggcagtggtg tctggagacc tggatttgag tcttggtgct atcaatcacc gtctgtgtt
2521 gagcaaggca tttggctgct gtaagcttat tgcttcatct gtaagcggtg gtttgtaatt
2581 cctgatcttc ccacctcaca gtgatgttgt ggggatccag tgagatagaa tacatgtaag
2641 tgtggttttg taatttaaaa agtgctatac taagggaag aattgaggaa ttaactgcat
2701' acgttttggt gttgcttttc aaatgtttga aaacaaaaaa aatgttaaga aatgggtttc
2761 ttgccttaac cagtctctca agtgatgaga cagtgaagta aaattgagtg cactaaacaa
2821 ataagattct gaggaagtct tatcttctgc agtgagtatg gcccgatgct ttctgtggct
2881 aaacagatgt aatgggaaga aataaaagcc tacgtgttgg taaatccaac agcaaggag
2941 atttttgaat cataataact cataagggtc tatctgttca gtgatgcctt cagagctctt
3001 gctgttagct ggcagctgac gctgctagga tagttagttt ggaaatggta cttcataata
3061 aactacacaa ggaaagtcag ccactgtgtc ttatgaggaa ttggacctaa taaattttag
3121 tgtgccttcc aaacctgaga atatatgctt ttggaagtta aaatttaaat ggcttttgcc
3181 acatacatag atcttcatga tgtgtgagtg taattccatg tggatatcag ttaccaaaca
3241 ttacaaaaaa attttatggc ccaaatgac caacgaaatt gttacaatag aatttatcca
3301 attttgatct ttttatattc ttctaccaca cctggaaaca gaccaataga cattttggg
3361 ttttataata ggaatttgta taaagcatta ctctttttca ataaattgtt ttttaattta
3421 aaaaaaggaa aaaaaaaaaa aaaaa (SEQ ID NO: 79)

```

FIGURE 43A

68/115

Claudin 1 (NM_021101)

MANAGLQLLGFI LAFLGWIGAIVSTALPQWRIYSYAGDNIVTAQ

AMYEGLWMSCVSQSTGQIQCKVFDLLNLSSTLQATRALMVVGILLGVIAIFVATVGM

KCMKCLEDDEVQKMRMAVIGGAIFLLAGLAILVATAWYGNRIVQEFYDPMTFVNARYE

FGQALFTGWAAAASLCLLGALLCCSCPRKTTSYPTPRPYPKPAPSSGKD YV (SEQ ID NO:80)

FIGURE 43B

69/115

Claudin 14 (NM_012130)

```

1  gtttgcttca ccttctgcca ggattgtaag tttcctgagg cctccccagt cctgcggaac
61  tggctccggc tggcacctga ggagcggcgt gaccccgagg gcccagggag ctgcccggct
121 ggcctaggca ggcagccgca ccatggccag caccggcgtg cagcttctgg gcttcctgct
181 cagcttcctg ggcattggtg gcacgttgat caccaccatc ctgccgcact ggcggaggac
241 agcgcacgtg ggcaccaaca tcctcacggc cgtgtcctac ctgaaagggc tctggatgga
301 gtgtgtgtgg cacagcacag gcatctacca gtgccagatc taccgatccc tgctggcgct
361 gcccacaagac ctccaggctg cccgcgccct catggtcac tctgcctgc tctcgggcat
421 agcctgcgcc tgcgccgtca tcgggatgaa gtgcaecgc tcgcccaagg gcacaccgc
481 caagaccacc tttgccatcc tcggcggcac cctcttcac ctggccggcc tctgtgcat
541 ggtggccgtc tcctggacca ccaacgacgt ggtgcagaac ttctacaacc cgctgctgcc
601 cagcggcatg aagtttgaga ttggccaggc cctgtacctg ggcttcatct cctcgtccct
661 ctgctcatt ggtggcacc tgctttgct gtctgccag gacgaggcac cctacaggcc
721 ctaccaggcc ccgccaggg ccaccacgac cactgcaaac accgcacctg cctaccagcc
781 accagctgcc taaaaagaca atcgggcccc ctcagtgacc tcggccacgc acagcgggta
841 caggctgaac gactacgtgt gagtccccac agcctgcttc tcccctgggc tgctgtgggc
901 tgggtccccg gcgggactgt caatggaggc aggggttcca gcacaaagtt tacttctggg
961 caatTTTTGT atccaaggaa ataatgtgaa tgcgaggaaa tgtctttaga gcacagggac
1021 agaggggggaa ataagaggag gagaaagctc tctataccaa agactgaaaa aaaaaatcct
1081 gtctgttttt gtatttatta tatatattta tgtgggtgat ttgataacaa gtttaataata
1141 aagtgaacttg ggagtttggc cagtggggtt ggtttgtgat ccaggaataa accttgcgga
1201 tgtggctgtt tatgaaaaaa aaaaaaaaaa aaa (SEQ ID NO:81)

```

FIGURE 44A

70/115

Claudin 14 (NM_012130)

MASTAVQLLGFLLSFLGMVGTLLITTLPHWRRTAHVGTNILTAV
SYLKGLWMECVWHSTGIYQCQIYRSLALPQDLQAARALMVISCLLSGIACACAVIGM
KCTRCAGTTPAKTTFAILGGTLFILAGLLCMVAVSWTTNDVVQNFYNPLLPSGMKFBI
GQALYLGFISSSLIGGTLLCLSCQDEAPYRPYQAPPRATTTTANTAPAYQPPAAYK
DNRAPSVTSATHSGYRLNDYV (SEQ ID NO:82)

FIGURE 44B

71/115

Tenascin-R (NM_003285)

```

1  ccttggtttc cgttgccagat tcccacaact ccatgctgtg tgctgcaggc tggctcctgaa
61  cccagatctc tggctgagag gatgggggca gatgggggaaa cagtggttct gaagaacatg
121 ctcatggcg tcaacctgat ccttctgggc tccatgatca agccttcaga gtgtcagctg
181 gaggtcacca cagaaagggg ccagagacag tcagtggagg aggaggagg cattgccaac
241 tacaacacgt ccagcaaaga gcagcctgtg gtcttcaacc acgtgtacaa cattaacgtg
301 cccttggaaca acctctgctc ctccaggcta gaggcctctg ctgagcagga ggtgagtga
361 gaagacgaga ctctggcaga gtacatgggc cagacctcag accacgagag ccaggtcacc
421 tttacacaca ggatcaactt ccccaaaaag gcctgtccat gtgccagttc agcccagggtg
481 ctgcaggagc tgctgagccg gatcgagatg ctggagaggg aggtgtcggg gctgcgagac
541 cagtgcacac ccaactgctg ccaagaaagt gctgccacag gacaactgga ctatatccct
601 cactgcagtg gccacggcaa ctttagcttt gagtcctgtg gctgcatctg caacgaaggc
661 tgggtttggca agaattgctc ggagccctac tgcccgtggg gttgctccag cgggggggtg
721 tgtgtggatg gccagtgcag ctctgggggg ctctgctggg acggggagtg tgtctgtgaa
781 cggtgcccaa cagactgcag ctcccggggg ctctgctggg acggggagtg tgtctgtgaa
841 gagccctaca ctggcgagga ctgcagggaa ctgaggtgcc ctggggactg ttcggggag
901 gggagatgtg ccaacgggtac ctgtttatgc gaggagggct acgttgggtg ggactgcggc
961 cagcggcagt gtctgaatgc ctgcagtggg cgaggacaat gtgaggaggg gctctgcgtc
1021 tgtgaagagg gctaccaggg ccctgactgc tcagcagttg cccctccaga ggacttgcca
1081 gtggctggta tcagcgacag gtccattgag ctggaatggg acggggccgag ggcagtgcag
1141 gaatatgtga tctcttacca gccgacggcc ctgggggggg tccagctcca gcagcgggtg
1201 cctggagatt ggagtgggtg caccatcacg gagctggagc caggtctcac ctacaacatc
1261 agcgtctacg ctgtcattag caacatctc agccttccca tctactgcaa ggtggccacc
1321 catctctcca ctctcaagg gctacaattt aagacgatca cagagaccac cgtggaggtg
1381 cagtgggagc ccttctcatt ttccttcgat ggggtgggaa tcagcttcat tccaaagaa
1441 aatgaagggg gagtgattgc tcaggtcccc agcgatgtta cgtcctttaa ccagacagga
1501 ctaaagcctg gggaggaata cattgtcaat gtgggtggctc tgaaagaaca ggcccgagc
1561 cccctacct cggccagcgt ctccacagtc attgacggcc ccacgcagat cctgggtcgc
1621 gatgtctcgg acaccgtggc ttttgtggag tggattcccc ctcgagccaa agtcgatttc
1681 attccttttg aatatggcct ggtgggcggg gaaggtggga ggaccacctt cgggctgcag
1741 cctccccctg gccaatctc agtgaggcc ctgcccgtg gctcccgata cgaggtgtca
1801 gtcagtgcg tccgagggac caacgagagc gattctgcca ccactcagtt cacaacagag
1861 atcgatgccc ccaagaactt gcgagttggt tctgcacag caaccagcct tgacctcgag
1921 tgggataaca gtgaagccga agttcaggag tacaagggtg tgtacagcac cctggcgggt
1981 gagcaatata atgaggtact ggtccccagg ggcattgggt caaccaccag ggccaccctg
2041 acagatctgg tacctggcac tgagtatgga gttggaatat ctgccgtcat gaactcacag
2101 caaagcgtgc cagccaccat gaatgccagg actgaacttg acagtcccc agacctcatg
2161 gtgacagcct cctcgggagc ctccatctcc ctcatctgga ccaaggccag tggccccatt
2221 gaccactacc gaattacctt taccocatcc tctgggattg cctcagaagt caccgtacct
2281 aaggacagga cctcatacac actaacagat ctagagcctg gggcagagta catcatttcc
2341 gtcactgctg agaggggtcg gcagcagagc ttggagtcca ctgtggatgc tttcacaggc
2401 ttccgtccca tctctcatct gcacttttct catgtgacct cctccagtgt gaacatcact
2461 tggagtgate catctcccc agcagacaga ctcatcttta actacagccc cagggatgag
2521 gaggaagaga tgatggaggt ctccctggat gccaccaaga ggcattgctg cctgatgggc
2581 ctgcaaccag ccacagagta tatttgtaac cttgtggctg tccatggcac agtgacctct
2641 gagccattg tgggctccat caccacagga attgatcccc caaaagacat cacaatttag
2701 aatgtgacca aggactcagt gatggtctcc tggagccctc ctggtgcac tttcgattac
2761 taccgagtat catatcgacc cacccaagtg ggacgactag acagctcagt ggtgcccac
2821 actgtgacag aattcaccat caccagactg aaccagcta ccgaatacga aatcagcctc
2881 aacagcgtgc ggggcaggga ggaaagcgag cgcactgtga ctctgtgca cacagccatg
2941 gacaaccctg tggatctgat tgctaccaat tctactcaa cagaagccct gctgcagtgg
3001 aaggcaccag tgggtgaggt ggagaactac gtcattgttc ttacacactt tgcagtcgt
3061 ggagagacca tccttggtga cggagtcaat gaggaatttc ggcttggtga cctgcttctc
3121 agcaccctc atactgccac catgtatgcc accaatggac ctctcaccag tggcaccatc
3181 agcaccact tttctactct cctggaccct ccggcaaacc tgacagccag tgaagtcacc
3241 agacaaagtg ccctgatctc ctggcagcct ccaggggcag agattgaaaa ttatgtcttg
3301 acctacaaat ccaccgacgg aagccgcaag gagctgattg tggatgcaga agacacctgg
3361 attcgactgg agggcctggt ggagaacaca gactacagc tgctcctgca ggcagcacag

```

FIGURE 45A

72/115

```

3421 gacaccacgt ggagcagcat cacctccacc gctttcacca caggaggccg ggtgttccct
3481 catccccaag actgtgcca gcatttgatg aatggagaca ctttgagtgg ggtttaccoc
3541 atcttctca atggggagct gagccagaaa ttacaagtgt actgtgatat gaccaccgac
3601 gggggcggtt ggattgtatt ccagaggcgg cagaatggcc aaactgattt tttccgaaa
3661 tgggctgatt accgtgttgg cttcgggaac gtggaggatg agttctggct ggggctggac
3721 aatatacaca ggatcacatc ccagggcgc tatgagctgc gcgtggacat gcgggatggc
3781 caggaggccg cttcgcctc ctacgacagg ttctctgtcg aggacagcag aaacctgtac
3841 aaactccgca taggaagcta caacggcact gcgggggact ccctcagcta tcatcaagga
3901 cgccctttct ccacagagga tagagacaat gatgttgag tgactaactg tgccatgtcg
3961 tacaaggag catggtggtg taagaactgc caccggacca acctcaatgg gaagtacggg
4021 gagtccaggc acagtcaggg catcaactgg taccattgga aaggccatga gttctccatc
4081 ccctttgtgg aaatgaagat gcgcccctac aaccaccgtc tcatggcagg gagaaaacgg
4141 cagtccttac agttctgagc agtgggcggc tgcaagccaa ccaatatttt ctgtcatttg
4201 tttgtatfff ataatatgaa acaagggggg agggtaatag caatgtgttt tgcaacatat
4261 taagagtatg tgaaggaagc agggatgtcg caggaatccg ctggctaaca tctgtctctg
4321 gtttctgctg ccctggagcc tgaccctcag tctccattct ccctcctacc caggcctcct
4381 caaccttcac ctcccttccc accaaggagg agaagtagga agttttctta aaggccaat
4441 tcaaagccaa gtcgtggggt gcagattgtt atggtgacag gcacacacat ttttctaccc
4501 ttcttctgag atgtcctctg ccttcagggt atttgtgatt ttgtcacagc ctgacatggc
4561 caggttctca cactggccca gagaaaagag cctcagcaag agagttttgc caacaattcc
4621 ccttaaaagg aaacagatca actacaccgc atcccaacaa ccaggttct tttccttct
4681 tccttccttc ctcccttctc tctttcctgc ctccc (SEQ ID NO:83)

```

FIGURE 45B

73/115

Tenascin-R (NM_003285)

MGADGETVVLKNMLIGVNLILLGSMIKPSECQLEVTTERVQRQS
VEEEGGIANYNTSSKEQPVVFNHVYNINVPLDNLCSGLEASAEQEVSAEDETIAEYM
GQTS DHESQVTFTHRINFPPKKACPCASSAQVLQELLSRIEMLEREVSVLRDQC NANCC
QESAATGQLDYIPHCSGHGNFSFESCGCI CNEGWF GKNCSEPYCPLGCSSRGVCVDGQ
CICDSEYSGDDCSELRCPTDCSSRGLCVDGECVCEEPTGTEDCRELRCPGD CSGKGR C
ANGTCLCEEGYVGEDCGQRQCLNACSGRGQCEEGLCVCEEYQGPDCSAVAPPEDLRV
AGISDRSIELEWDGPM AVTEYVISYQPTALGGLQLQQRVPGDWSGVTITELPGLTYN
ISVYAVISNILSLPITAKVATHLSTPQGLQFKTITETTVEVQWEPFSFSFDGWEISFI
PKNNEGGVIAQVPSDVT SFNQ TGLKPGE EYIVNVVALKEQARSPPTSASVSTVIDGPT
QILVRDVSDTVAFVEWIPPRAKVDFILLKYGLVGGEGGRTTFR LQPPLSQYSVQALRP
GSRYEVS VSAVRGTNESDSATTQFTTEIDAPKNLRVGSRTATSLDLEWDNSEAEVQEY
KVVYSTLAGEQYHEVLVPRGIGPTTRATLTDLVPGTEYGVGISAVMNSQQSVPATMNA
RTELDSPRDLMTASSET SISLIWTKASGPIDHYRITFTPSSG IASEVTVPKDRTSYT
LTDLEPGA EYIISVTAERGRQQSLESTVDAFTGFRPISHLHF SHVTSSSVNITWSDPS
PPADRLILNYSRDEEEEMMEVSLDATKRHAVLMGLQPATEYIVNLVAVHGTVTSEPI
VGSITTGIDPPKDITISNVTKDSVMVSWSPPVASF DYYRVSYRPTQVGR LDSSVVPNT
VTEFTITRLNPATEYEISLNSVRGREESERIC TLVHTAMDNPNVDLIATNITPTEALLQ
WKAPVGEVENYVIVLTHFAVAGETILVDGVSEEFRLVDLLPSTHYTATMYATNGPLTS
GTISTNFS TLLDPPANLTASEVTRQSALISWQPPRAE IENYVLT YKSTDGSRKELIVD
AEDTWIRLEGLLENTDYTVLLQAAQDTTWSSITSTAFTTGGRVFPHPQDCAQHLMNGD
TLSGVYPIFLNGELS QKLQVYCDMTTDGGGWIVFQRRQNGQTDFFRKWADYRVGFGNV
EDEFWLGLDNIHRITSQGRYELRVDMRDGEAAAFASYDRFSVEDSRNLYKLRI GSYNG
TAGDSL SYHQGRPFSTEDRDNDVAVTNCAMSYKGAWWYKNCHRTNLNGKYGESRHSQG
INWYHWKGHEFSIPFVEMKMRPYNHRLMAGRKRQSLQF (SEQ ID NO:84)

FIGURE 45C

74/115

CAD3 (NM-001793)

```

1 aaagggggcaa gagctgagcg gaacaccggc ccgcccgtcg ggcagctgct tcacccctct
61 ctctgcagcc atggggctcc ctctgaggacc tctcgcgtct ctccctcttc tccagggttg
121 ctggctgcag tgcgcggcct ccgagccgtg ccgggcggtc ttcaggaggg ctgaagtgc
181 ctggaggcg ggaggcgcg agcaggagcc cggccaggcg ctggggaaag tattcatggg
241 ctgccctggg caagagccag ctctgtttag cactgataat gatgacttca ctgtgcggaa
301 tggcgagaca gtccaggaaa gaaggtcact gaaggaaagg aatccattga agatcttccc
361 atccaaacgt atcttacgaa gacacaagag agattgggtg gttgctccaa tatctgtccc
421 tgaaaatggc aagggtccct tccccagag actgaatcag ctcaagtcta ataaagatag
481 agacaccaag attttctaca gcatcacggg gccgggggca gacaggggtg
541 cttcgctgta gagaaggaga caggctggtt gttgttgaaat aagccactgg accgggagga
601 gattgccaag tatgagctct ttggccacgc tgtgtcagag aatgggtgct cagtggagga
661 ccccatgaac atctccatca tctgaccga ccagaatgac cacaagccca agtttaccca
721 ggacaccttc cgaggagtg tcttagaggg agtctacca ggtacttctg tgatgcaggt
781 gacagccacg gatgaggatg atgccatcta cactacaaat ggggtgggtg cttactccat
841 ccatagccaa gaaccaaagg acccacacga cctcatgttc accattcacc ggagcacagg
901 caccatcagc gtcatctcca gtggcctgga ccgggaaaaa gtccctgagt acacactgac
961 catccaggcc acagacatgg atggggacgg ctccaccacc acggcagtgg cagtagtgga
1021 gatccttgat gccaatgaca atgctcccat gtttgacccc cagaagtacg agggccatgt
1081 ccctgagaat gcagtgggcc atgaggtgca gaggctgacg gtccactgatc tggacgcccc
1141 caactcacca gcgtggcggt ccacctacct tatcatgggc ggtgacgacg gggaccattt
1201 taccatcacc acccacctg agagcaacca gggcatcctg acaaccagga agggtttgga
1261 ttttgaggcc aaaaaccagc acacctgta cgttgaagtg accaacgagg ccccttttgt
1321 gctgaagctc ccaacctcca cagccaccat agtgggtccac gtggaggatg tgaatgaggc
1381 acctgtgttt gtcccacct ccaaagtcgt tgagggtccag gagggtcatc cactgggga
1441 gcctgtgtgt gtctacactg cagaagacc tgacaaggag aatcaaaaga tcagctaccg
1501 catcctgaga gaccagcag ggtggctagc catggacca gacagtgggc aggtcacagc
1561 tgtgggcacc ctgcaccgtg aggatgagca gtttgtgagg aacaacatct atgaagtcac
1621 ggtcttgccc atggacaatg gaagccctcc caccactggc acgggaaccc ttctgctaac
1681 actgattgat gtcaatgacc atggcccagt cctgagccc cgtcagatca ccatctgcaa
1741 ccaaagccct gtgcgccagg tgctgaacat cacggacaag gacctgtctc cccacacctc
1801 ccctttccag gccagctca cagatgactc agacatctac tggacggcag aggtcaacga
1861 ggaagggtgac acagtgggtc tgtccctgaa gaagttcctg aagcaggata catatgacgt
1921 gcacctttct ctgtctgacc atggcaacaa agagcagctg acggtgatca gggccactgt
1981 gtgcgaactgc catggccatg tcgaaacctg ccctggaccc tggaaaggag gtttcatcct
2041 ccctgtgctg ggggctgtcc tggctctgct gttcctcctg ctggtgctgc cttgttggt
2101 ggaagaagag cggaagatca aggagccct cctactccca gaagatgaca cccgtgacaa
2161 cgtcttctac tatggcgaag aggggggtgg cgaagaggac caggactatg acatcaccca
2221 gctccaccga ggtctggagg ccaggccgga ggtggttctc cgcaatgacg tggcaccac
2281 catcatcccg acacccatgt accgtcctcg gccagccaac ccagatgaaa tcggcaactt
2341 tataattgag aacctgaagg cggctaacac agaccccaca gccccgccct acgacacctc
2401 cttggtgttc gactatgagg gcagcggctc cgacgcgcgc tccctgagct ccctcacctc
2461 ctccgcctcc gaccaagacc aagattacga ttatctgaac gagtggggca gccgctcaa
2521 gaagctggca gacatgtacg gtggcgggga ggacgactag gcggcctgcc tgcagggtg
2581 gggaccaaac gtcaggccac agagcatctc caaggggtct cagtcccccc ttcagctgag
2641 gacttcggag cttgtcagga agtggccgta gcaacttggc ggagacaggc tatgagtctg
2701 acgttagagt gggtgcttcc ttgaccttcc aggatggagg aatgtgggca gtttgacttc
2761 agcactgaaa acctctccac ctgggccagg gttgcctcag aggccaaagt tccagaagcc
2821 tottacctgc cgtaaaatgc tcaacctgt gtctgggccc tgggcctgct gtgactgacc
2881 tacagtggac tttctctctg gaatggaacc ttcttaggac tctggtgca acttaatttt
2941 tttttttaat gctatcttca aaacgttaga gaaagttctt caaaagtgca gccagagct
3001 gctgggcccc ctggccgtcc tgcatttctg gtttccagac cccaatgcct cccattcgga
3061 tggatctctg cgtttttata ctgagtgtgc ctaggttgcc ccttattttt tattttccct
3121 gttgcgttgc tatagatgaa gggtaggac aatcgtgtat atgtactaga acttttttat
3181 taaagaaact tttcccagaa aaaaa (SEQ ID NO:85)

```

FIGURE 46A

75/115

CAD3 (NM-001793)

MGLPRGPLASLLLLQVCWLQCAASEPCRAVFREAEVTLEAGGAE
QEPGQALGKVFMGCPGQEPALFSTDNDFFTNRGETVQERRSLKERNPLKIFPSKRIL
RRHKRDWVAPISVPENGKGPFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAV
EKETGWLLLLNKPLDREEIAKYELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQD
TFRGSVLEGVLPGTSMQVTATDEDDAIYTYNGVVAYSIIHSQEPKDPHDLMTIHRST
GTISVISSGLDREKVPEYTLTIQATMDGDGSTTTAVAVVEILDANDNAPMFDPOKYE
AHVPENAVGHEVQRLTVTDLDAPNSPAWRATYLMGGDDGDHFTITTHPESNQGILTT
RKGLDFEAKNQHTLYVEVTNEAPFVLKLPTSTATIVVHVEDVNEAPVFVPPSKVVEVQ
EGIPTGEPVCVYTAEDDPKENQKISYRILRDPAGWLAMDPSGQVTAVGTLTREDEQF
VRNNIYEVMLAMDNGSPPTTGTGTLTLIDVNDHGPVPEPRQITICNQSPVRQVLN
ITDKDLSPHTSPFQAQLTDDSDIYWTAEVNEEGDTVVLSLKKFLKQDTYDVHLSLSDH
GNKEQLTVIRATVCDCHGHVETCPGPWKGGFILPVLGAVLALLFLLLVLVVLRKKRK
IKEPLLLPEDDTRDNVFIYGEEGGEEDQDYDITQLHRGLEARPEVVLNRNDVAPTIIP
TPMYRPRPANPDEIGNFIIENLKAANTDPTAPPYDTLLVFDYEGSGSDAASLSSLTSS
ASDQDQDYDYLNEWGSRFKKLADMYGGGEDD (SEQ ID NO:86)

FIGURE 46B

76/115

CONT (NM_001843)

```

1 gctgtgccgc accgaggcga gcaggagcag ggaacagggtg tttaaaatta tccaactgcc
61 atagagctaa attctttttt ggaaaattga accgaacttc tactgaatac aagatgaaaa
121 tgtggttgct ggtcagtcac cttgtgataa tatctattac tacctgttta gcagagttta
181 catggtatag aagatatggt catggagttt ctgaggaaga caaaggattt ggaccaattt
241 ttgaagagca gccaatcaat accatttatc cagaggaatc actggaagga aaagtctcac
301 tcaactgtag ggcacgagcc agccctttcc cggtttacia atggagaatg aataatgggg
361 acgttgatct cacaagtgat cgatacagta tggtaggagg aaacctgtt atcaacaacc
421 ctgacaaaca gaaagatgct ggaatatact actgttttagc atctaataac tacgggatgg
481 tcagaagcac tgaagcaacc ctgagctttg gatattctga tcccttccca cctgagggaac
541 gtcctgaggt cagagtaaaa gaagggaag gaattggtgct tctctgtgac cccccatacc
601 attttccaga tgatcttagc tatcgctggc ttctaaatga atttctgtta tttatcacaa
661 tggataaacg gcgatttgtg tctcagacaa atggcaatct ctacattgca aatgttgagg
721 cttccgacaa aggcaattat tctgtctttg tttccagtc tctctattaca aagagcgtgt
781 tcagcaaatt catcccactc attccaatac ctgaacgaac aacaaaacca tatcctgctg
841 atattgtagt tcagttcaag gatgtatatg cattgatggg ccaaaatgtg accttagaat
901 gttttgcact tggaaatcct gttccggata tccgatggcg gaaggttcta gaaccaatgc
961 caagcactgc tgagattagc acctctgggg ctgttcttaa gatcttcaat attcagctag
1021 aagatgaagg catctatgaa tgtgaggctg agaacattag aggaaaggat aaacatcaag
1081 caagaattta tgttcaagca ttccctgagt gggtagaaca catcaatgac acagaggtgg
1141 acataggcag tgatctctac tggccttgtg tggccacagg aaagcccatc cctacaatcc
1201 gatggttgaa aaatggatat gcgtatcata aaggggaatt aagactgtat gatgtgactt
1261 ttgaaaatgc cggaatgtat cagtgcatac ctgaaaacac atatggagcc atttatgcaa
1321 atgctgagtt gaagatcttg gcgttggctc caacttttga aatgaatcct atgaagaaaa
1381 agatcctggc tgctaaaggt ggaagggtga taattgaatg caaacctaa gctgcaccga
1441 aaccaaagtt tcatggagt aaaggacag agtggcttgt caatagcagc agaatactca
1501 tttgggaaga tggtagcttg gaaatcaaca acattacaag gaatgatgga ggtatctata
1561 catgctttgc agaaaataac agagggaag ctaatagcac tggaaacctt gttatcacag
1621 atcctacgag aattatattg gcccgaatta atgccgatat cacagttgga gaaaacgcca
1681 ccatgcagtg tgctgcgtcc tttgatcctg ccttggatct cacatttgtt tggctcctca
1741 atggctatgt gatcgatttt aacaaagaga atattcacta ccagaggaat tttatgctgg
1801 attccaatgg ggaattacta atccgaaatg cgcagctgaa acatgctgga agatacacat
1861 gcaactgccc gacaattgtg gacaattctt cagcttcagc tgaccttgta gtgagaggcc
1921 ctccaggccc tccagggtgt ctgagaatag aagacattag agccacttct gtggcactta
1981 cttggagcgg tggttcagac aatcatagtc ctatttctaa atacactatc cagaccaaga
2041 ctattctttc agatgactgg aaagatgcaa agacagatcc cccaattatt gaaggaaata
2101 tggaggcagc aagagcagtg gacttaatcc catggatgga gtatgaattc cgcgtggtag
2161 caaccaatac actgggtaga ggagagccca gtataccatc taacagaatt aaaacagacg
2221 gtgctgcacc aaatgtggct ccttcagatg taggaggtgg aggtggaaga aacagagagc
2281 tgaccataac atgggcgcct ttgtcaagag aataccacta tggcaacaat tttggttaca
2341 tagtggcatt taagccattt gatggagaag aatggaaaaa agtcacagtt actaatcctg
2401 atactggcgg atatgtccat aaagatgaaa ccatgagccc ttcactgca tttcaagtta
2461 aagtcaaggc cttcaacaac aaaggagatg gaccttacag cctagtagca gtcattaatt
2521 cagcacaaga cgctcccagt gaagcccaa cagaagtagg tgtaaaagtc ttatcatctt
2581 ctgagatata tgttcattgg gaacatgttt tagaaaaaat agtggaaagc tatcagattc
2641 ggtattgggc tgcccatgac aaagaagaag ctgcaaacag agttcaagtc accagccaag
2701 agtactcggc caggctcgag aaccttctgc cagacacca gtattttata gaagtcgggg
2761 cctgcaatag tgcagggtgt ggacctcaa gtgacatgat tgaggctttc accaagaaag
2821 cacctcctag ccagcctcca aggatcatca gttcagtaag gtctggttca cgctatataa
2881 tcacctggga tcatgtcgtt gcactatcaa atgaatctac agtgacggga tataagggtac
2941 tctacagacc tgatggccag catgatggca agctgtattc aactcacaaa cactccatag
3001 aagtcccaat cccagagat ggagaatacg ttgtggaggt tcgcgcgcac agtgatggag
3061 gagatggagt ggtgtctcaa gtcaaaattt cagggtgcacc caccctatcc ccaagtcttc
3121 tcggcttact gctgcctgcc tttggcatcc ttgtctactt ggaattctga atgtgttgtg
3181 acagctgctg ttcccatccc agctcagaag acacccttca accctgggat gaccacaatt
3241 ccttccaatt tctgcggctc catcctaagc caaataaatt atactttaac aaactattca
3301 actgatttac aacacacatg atgactgagg cattcgggaa ccccttcatc caaagaata
3361 aactttttaa tggatataaa tgatttttaa ctcgttccaa tatgccttat aaaccactta
3421 acctgat (SEQ ID NO:87)

```

FIGURE 47A

77/115

CONT (NM_001843)

MKMWLLVSHLVIIISITTCLEFTWYRRYGHGVSEEDKGFPIPE
EQPINTIYPEESLEGKVSINCRARASFPVYKWRMNNGDVDLTSDRYSMVGGNLVINN
PDKQKDAGIYYCLASNNGMVRSTEATLSFGYLDPPPEERPEVRVKEGKGMVLLCDP
PYHFPDDL SYRWLLNEFPVFITMDKRRFVSQTNGNLYIANVEASDKGNYS CFVSSPSI
TKSVFSKFIPLIPERTTKPYPADIVVQFKDVYALMGQNVTL ECFALGNPVPDIRWR
KVLEPMPSTAEISTSGAVLKI FNIQLEDEGIYECEAENIRGKDKHQARIYVQAFPEWV
EHINDTEVDIGSDLYWPCVATGKPIPTIRWLKNGYAYHKGELRLYDVT FENAGMYQCI
AENTYGAIYANAELKILALAPT FEMNPMKKKILAAGGRVIECKPKAAPKPKFSWSK
GTEWLVNSSRILIWEDGSLEINNITRNDGGIYTCFAENNRGKANSTGTLVITDPTRII
LAPINADITVGENATMQCAASFDPALDLTFVWSFNGYVIDFNKENIHYQRNFMLDSNG
ELLIRNAQLKHAGRYTCTAQTIVDNSSASADLVVRGPPGPPGGLRIEDIRATSVALTW
SRGSDNHSPISKYTIQTKTILSDDWKDAKTDPPII EGNMEAARAVDLIPWMEYEFVRV
ATNTLGRGEP SIPS NR IKTDGAAPNVAPSDVGGGGGRNREL TITWAPLSREYHYGNF
GYIVAFKPF DGE EWKKVTVTNPDTGRYVHKDET MSPSTAFQVKVKAFNNKGDGPYSLV
AVINSAQDAPSEAPTEVG VKVLSSSEISVHWEHVLEKIVESYQIRYWAAHDKEEAANR
VQVTSQEYSARLENLLPDTQYFIEVGACNSAGCGPPSDMIEAFTKKAPPSQPPRIISS
VRSGSRYIITWDHVVALSNESTVTGYKVL YRPDGHGKLYSTHKHSIEVP IPRDGEY
VVEVRAHSDGGDGVVSQVKISGAPTLSPSLLGLLLPAFGILVYLEF (SEQ ID NO:88)

FIGURE 47B

78/115

Osteopontin (NM_000582)

```

1  ctccctgtgt  tgggtggagga  tgtctgcagc  agcattttaa  ttctgggagg  gcttggttgt
61  cagcagcagc  aggaggaggc  agagcacagc  atcgtcggga  ccagactcgt  ctcaggccag
121  ttgcagcctt  ctcagccaaa  cgccgaccaa  ggaaaactca  ctaccatgag  aattgcagtg
181  atttgctttt  gcctcctagg  catcacctgt  gccataccag  ttaaacaggc  tgattctgga
241  agttctgagg  aaaagcagct  ttacaacaaa  taccagatg  ctgtggccac  atggctaaac
301  cctgacccat  ctcagaagca  gaatctccta  gccccacaga  cccttccaag  taagtccaac
361  gaaagccatg  accacatgga  tgatatggat  gatgaagatg  atgatgacca  tgtggacagc
421  caggactcca  ttgactcgaa  cgactctgat  gatgtagatg  aactgatga  ttctcaccag
481  tctgatgagt  ctcaccattc  tgatgaatct  gatgaactgg  tctactgatt  tcccacggac
541  ctgccagcaa  ccgaagtttt  cactccagtt  gtccccacag  tagacacata  tgatggccga
601  ggtgatagt  tggtttatgg  actgagggtc  aaatctaaga  agtttcgcag  acctgacatc
661  cagtaccctg  atgctacaga  cgaggacatc  acctcacaca  tggaaagcga  ggagttgaat
721  ggtgcataca  aggccatccc  cggtgcccag  gacctgaacg  cgcttctga  ttgggacagc
781  cgtgggaagg  acagttatga  aacgagtcag  ctggatgacc  agagtgctga  aaccacagc
841  cacaagcagt  ccagattata  taagcggaaa  gccaatgatg  agagcaatga  gcattccgat
901  gtgattgata  gtcaggaact  ttccaaagtc  agccgtgaat  tccacagcca  tgaatttcac
961  agccatgaag  atatgctggg  tgtagacccc  aaaagtaagg  aagaagataa  acacctgaaa
1021  tttcgtattt  ctcatgaatt  agatagtgca  tcttctgagg  tcaattaaaa  ggagaaaaaa
1081  tacaatttct  cactttgcat  ttagtcaaaa  gaaaaaatgc  tttatagcaa  aatgaaagag
1141  aacatgaaat  gcttctttct  cagtttattg  gttgaatgtg  tatctatttg  agtctggaaa
1201  taactaatgt  gtttgataat  tagtttagtt  tgtggcttca  tggaaactcc  ctgtaaacta
1261  aaagcttcag  ggttatgtct  atgttcattc  tatagaagaa  atgcaaacta  tctactgtatt
1321  ttaatatatt  ttattctctc  atgaatagaa  atttatgtag  aagcaaacaa  aatactttta
1381  cccacttaaa  aagagaatat  aacattttat  gtcactataa  tcttttgttt  ttaagttag
1441  tgtatatatt  gttgtgatta  tctttttgtg  gtgtgaataa  atcttttatc  ttgaatgtaa
1501  taagaatttg  gtggtgtcaa  ttgcttattt  gttttccac  ggttgtccag  caattaataa
1561  aacataacct  tttttactgc  ctaaaaaaaa  aaaaaaaaaa  aaaaaaaaaa  aaaaaa (SEQ
ID NO:89)

```

FIGURE 48A

79/115

Osteopontin (NM_000582)

MRIAVICFCLLGITCAIPVKQADSGSSEEKQLYNKYPDAVATWL
NPDPSQKQNLLAPQTLPSKSNESHDMDDMDEDDDDHVDSQDSIDSNDSDDVDDTDD
SHQSDESHHSDESDELVTDFPTDLPATEVFTPVVPTVDTYDGRGDSVVYGLRSKSKKF
RRPDIQYPDATDEDITSHMESEELNGAYKAI PVAQDLNAPSDWDSRGKDSYETSQLDD
QSAETHSHKQSRLYKRKANDESNEHSDVIDSQELSKVSREFHSHEFHSHEDMLVVDPK
SKEEDKHLKFRISHELDSASSEVN (SEQ ID NO:90)

FIGURE 48B

80/115

Galectin 8 (NM_006499)

```

1  tggacttggg  tccgaggcag  acgaggaagc  tgagaaaacc  ctggcggtga  ccccggtggac
61  ctgggcgccc  cggaaggtc  cagcgcttgg  tccaggcagg  cggggatgtg  cgggtgaccac
121  cctggtcctg  aaaagtccag  ccccggaatct  ccctccctcc  tagacctgga  ggcctggaac
181  agccagccgc  ccacggacgc  cagagccggg  aaccctgacg  gcacttagct  gctgacaaac
241  aacctgctcc  gtggacgcct  gaaacaccag  tctttggggc  cagtgcctca  gtttcaatcc
301  aggtaacctt  taaatgaaac  ttgcctaaaa  tcttaggtca  tacacagaag  agactccaat
361  cgacaagaag  ctggaaaaga  atgatgttgt  ccttaaacia  cctacagaat  atcatctata
421  acccggtaat  cccgtatgtt  ggcaccattc  ccgatcagct  ggatccctga  actttgattg
481  tgatatgtgg  gcatgttcct  agtgacgcag  acagattcca  ggtggatctg  cagaatggca
541  gcagtgtgaa  acctcgagcc  gatgtggcct  ttcatttcaa  tcctcgtttc  aaaagggccg
601  gctgcattgt  ttgcaatact  ttgataaatg  aaaaatgggg  acgggaagag  atcacctatg
661  acacgccttt  caaaagagaa  aagtcttttg  agatcgtgat  tatggtgcta  aaggacaaat
721  tccaggtggc  tgtaaatgga  aaacatactc  tgctctatgg  ccacaggatc  ggcccagaga
781  aaatagacac  tctgggcatt  tatggcaaag  tgaatattca  ctcaattggg  tttagcttca
841  gctcggactt  acaaagtacc  caagcatcta  gtctggaact  gacagagata  agtagagaaa
901  atgttccaaa  gtctggcacg  ccccgacttc  agactgtctc  tcctccctgg  gatttacagg
961  gtcattggctc  tgaaacattc  tgtagtgttc  tttggacacg  agttttcctg  gagatcgctt
1021  tctgcaggcc  tattggtctg  actgtggcct  cttttcagag  cctgccattc  gctgcaaggt
1081  tgaacacccc  catgggccct  ggacgaactg  tcgtcgttaa  aggagaagtg  aatgcaaagt
1141  ccaaaagctt  taatgttgac  ctactagcag  gaaaatcaaa  ggatattgct  ctacacttga
1201  acccacgcct  gaatattaaa  gcatttghta  gaaattcttt  tcttcaggag  tcctggggag
1261  aagaagagag  aaatattacc  tctttcccat  ttagtcctgg  gatgtacttt  gagatgataa
1321  tttactgtga  tgtagagaa  ttcaagggtg  cagtaaattg  cgtacacagc  ctggagtaca
1381  aacacagatt  taaagagctc  agcagtattg  acacgctgga  aattaatgga  gacatccact
1441  tactggaagt  aaggagctgg  tagcctacct  acacagctgc  taaaaaaccc  aaaatacaga
1501  atggcttctg  tgatactggc  cttgctgaaa  cgcactctac  tgtcatttca  ttgtttatat
1561  tgttaaaatg  agcttgtgca  ccattagatc  ctgctgggtg  ttctcagtc  ttgccatgaa
1621  gtatggtggt  gtctagcact  gaatggggaa  actgggggca  gcaacactta  tagccagtta
1681  aagccactct  gccctctctc  ctactttggc  tgactcttca  agaatgccat  tcaacaagta
1741  tttatggagt  acctactata  atacagtagc  taacatgtat  tgagcacaga  ttttttttgg
1801  taaaactgtg  aggagctagg  atatatactt  ggtgaaacaa  accagtatgt  tcctgttctt
1861  cttgagcttc  gactcttctg  tgctctattg  ctgcgcaactg  ctttttctac  aggcattaca
1921  tcaactccta  aggggtcctc  tgggattagt  taagcagcta  ttaaatacacc  cgaagacact
1981  aattttacaga  agacacaact  ccttccccag  tgatcaactgt  cataaccagt  gctctaccgt
2041  atcccatcac  tgaggactga  tgttgactga  catcatttta  tcgtaataaaa  catgtggctc
2101  tattagctgc  aagctttacc  aagtaattgg  catgacatct  gagcacagaa  attaaaggcaa
2161  aaaaccaaag  caaaacaaat  acatggtgct  gaaattaact  tgatgccaaag  cccaaggcag
2221  ctgatttctg  tgtatttgaa  cttaggggcaa  atcagagtct  acacagacgc  ctacagaaag
2281  tttcaggaag  aggcaagatg  cattcaattt  gaaagatatt  tatgggcaac  aaagtaaggt
2341  caggattaga  cttcaggcat  tcataaggca  ggcactatca  gaaagtgtac  gccaactaag
2401  ggaccacaaa  agcaggcaga  ggtaatgcag  aaatctgttt  tgttcccatg  aaatcaccaa
2461  tcaaggcctc  cgttcttcta  aagattagtc  catcatcatt  agcaactgag  atcaaagcac
2521  tcttccactt  tacgtgatta  aatcaaacc  tgtatcagca  aaaaaaaaaa  aaaaaaaaaa
2581  aaaaaaaaaa  aaa (SEQ ID NO:91)

```

FIGURE 49A

81/115

Galectin 8 (NM_006499)

MLSLNNLQNI IYNPVI PYVGTIPDQLDPGTLIVICGHVPSADR
FQVDLQNGSSVKPRADVAHFHFNPRFKRAGCIVCNTLINEKWGREEITYDTPFKREKSF
EIVIMVLKDKFQVAVNGKHTLLYGHRIGPEKIDTLGIYGKVNIHSIGFSFSSDLQSTQ
ASSLELTEISRENVPKSGTPQLQTVSPSWDLQGHGSETFCSVLWTRVFLEIAFCRPIG
LTVASFQSLPFAARLNTPMGPGRTVVVKGEVNANAKSFNVDLLAGKSKDIALHLNPRL
NIKAFVRNSFLQESWGEEERNITSFPFSPGMYFEMIIYCDVREFKVAVNGVHSLEYKH
RFKELSSIDTLEINGDIHLLEVRSW (SEQ ID NO:92)

FIGURE 49B

82/115

PGS1 (bihlycan, NM_001711)

```

1 agcctcccgc ccgcgcgcctc tgtctccctc tctccacaaa ctgcccagga gtgagtagct
61 gcttttcggtc cgccgggacac accgggacaga tagacgtgcg gacggcccccac cccccagcc
121 cgccaactag tcagcctgcg cctgggcgcct cccctctcca ggtccatccg ccatgtggcc
181 cctgtggcgc ctcgtgtctc tgttggccct gagccaggcc ctgccctttg agcagagagg
241 cttctgggac ttcaccctgg acgatgggccc attcatgatg aacgatgagg aagcttcggg
301 cgctgacacc tcgggcgcctc tggaccgcga ctctgtcaca cccacctaca gcgccatgtg
361 tccttttcggc tgccactgcc acctgcgggt ggttcagtgc tccgacctgg gtctgaagtc
421 tgtgccccaa gagatctccc ctgacaccac gctgctggac ctgcagaaca acgacatctc
481 cgagctccgc aaggatgact tcaagggtct ccagcacctc tacgccctcg tcctggtgaa
541 caacaagatc tccaagatcc atgagaaggc cttcagccca ctgcggaagc tgcagaagct
601 ctacatctcc aagaaccacc tgggtggagat cccgccccaa ctaccagct cctgggtgga
661 gctccgcac cagacaacc gcacccgcaa ggtgccccaa ggagtgttca gcgggctccg
721 gaacatgaac tgcacgaga tgggcgggaa cccactggag aacagtggct ttgaacctgg
781 agccttcgat ggcctgaagc tcaactacct gcgcacatca gagggcaagc tgactggcat
841 ccccaaagac ctccctgaga cctgaatga actccacctc gaccacaaca aaatccaggc
901 catcgaactg gaggacctgc ttgcctactc caagctgtac aggttgggccc tagggcaca
961 ccagatcagg atgatcgaga acgggagcct gagcttctctg cccaccctcc gggagctcca
1021 cttggacaac aacaagttgg ccagggtgcc ctcagggtct ccagacctca agctcctcca
1081 ggtggtctat ctgcactcca acaacatcac caaagtgggt gtcaacgact tctgtcccat
1141 gggcttcggg gtgaagcggg cctactacaa cggcatcagc ctcttcaaca acccctgcc
1201 ctactgggag gtgcagccgg ccactttccg ctgcgtcact gaccgcctgg ccatccagtt
1261 tggcaactac aaaaagtaga ggcagctgca gccaccgcgg ggcctcagtg ggggtctctg
1321 gggaacacag ccagacatcc tgatggggag gcagagccag gaagctaagc cagggccag
1381 ctgcgtccaa cccagcccc cactcgggt cctgacccc agctcgatgc cccatcaccg
1441 cctctccctg gctcccaagg gtgcaggtgg gcgcaaggcc cggcccccac cacatgttcc
1501 cttggcctca gagctgcccc tgctctccca ccacagccac ccagaggcac ccatgaagc
1561 ttttttctcg ttcactcca aacccaagt tccaaggctc cagtcctagg agaacagtcc
1621 ctgggtcagc agccaggagg cgggtccataa gaatggggac agtgggctct gccagggtg
1681 ccgcacctgt ccagacacac atgttctgtt cctcctcctc atgcatttcc agcctttcaa
1741 ccctcccga ctctgcggct cccctcagcc ccttgcaag ttcatggcct gtccctcca
1801 gaccctgct ccactggccc ttcgaccagt cctccctct gtctctctt tccccgtcct
1861 tcctctctct ctctctctct ctctctctct ctttctgtgt gtgtgtgtgt gtgtgtgtgt
1921 gtgtgtgtgt gtgtgtgtgt cttgtgcttc ctgagacctt tctcgcttct gagcttgggtg
1981 gcctgttccc tccatctctc cgaacctggc ttgcctgtc cctttcactc cacacctct
2041 ggccttctgc cttgagctgg gactgcttcc tgtctgtccg gcctgcacc agccccgtcc
2101 cacaaaacc cagggacagc ggtctcccca gcctgccctg ctcaggcctt gccccaaac
2161 ctgtactgtc ccggaggagg ttgggaggtg gaggccagc atcccgcga gatgacacca
2221 tcaaccgcca gactcccaga caccggtttt cctagaagcc cctcaccctc actggccac
2281 tgggtggetag gtctccctt atccttctgg tccagcgcaa ggaggggctg cttctgaggt
2341 cgggtggctgt ctttccatta aagaacacc gtgcaacgtg aaaaaaaaaa aaaaaaaaaa
2401 a (SEQ ID NO:93)

```

FIGURE 50A

83/115

PGS1 (bihlycan, NM_001711)

MWPLWRLVSLALSQLPFEQRGFWDFTLDDGPFMMNDEEASGA
DTSGVLDPDSVTPYTSAMCPFGCHCHLRVVQCSDLGLKSVPEISPD'TLLDLQNNDI
SELRKDDFKGLQHLYALVLVNNKISKIHEKAFSPLRKQLQKLYISKNHLEI'PPNLPSS
LVELRIHDNRIRKVPKGVFSGLRNMNCIEMGGNPLENSGFEPGAFDGLKLN'YLRISEA
KLTGIPKDLPETLNE'LDHNKIQAIELEDLLRYSKLYRLGLGHNQIRMIENGSL'SFL
PTLRELHLDNNKLARVPSGLPDLKLLQVVYLHSNNITKVG'VNDFCPMGFGVKRAY'YNG
ISLFNNPV'PYWEVQPATFRCVTDRLAIQFGNYKK (SEQ ID NO:94)

FIGURE 50B

84/115

Frizzled 2 (NM_001466)

```

1  cgagtaaagt ttgcaaagag ggcgaggagg cggcagccgc agcgaggagg cggcgaggaa
61  gaagcgcagt ctccgggttg ggggcggggg cggggggggc gccaggagc cgggtggggg
121  gcggcggcca gcatgcggcc ccgcagcgcc ctgccccgcc tgctgctgcc gctgctgctg
181  ctgcccgcgc ccgggcccgc ccagttccac ggggagaagg gcatctccat cccggaccac
241  ggcttctgcc agcccatctc catcccgcgt tgcacggaca tcgcctacaa ccagaccatc
301  atgcccacc ttctgggcca cacgaaccag gaggacgcag gcctagaggt gcaccagttc
361  tatccgctgg tgaagggtga gtgctcgccc gaactgcgct tcttcctgtg ctccatgtac
421  gcacccgtgt gcaccgtgct ggaacaggcc atcccgcgt gccgctctat ctgtgagcgc
481  gcgcgccagg gctgcgaagc cctcatgaac aagttcgggt ttcagtgggc cgagcgccgt
541  cgctgcgagc acttcccgcg ccacggcgcc gagcagatct gcgtcgggcc gaaccactcc
601  gaggacggag ctcccgcgt actcaccacc gcgcgcgcgc cgggactgca cgggggtgcc
661  gggggcaccc cgggtggccc gggcgggcgc ggcgctcccc cgcgctacgc cacgctggag
721  cacccttcc actgcccgcg cgtcctcaag gtgccatcct atctcagcta caagtttctg
781  ggcgagcgtg attgtgctgc gccctgcgaa cctgcgcggc ccgatgggtc catgttcttc
841  tcacaggagg agacgcgttt cgcgcgcctc tggatcctca cctggtcggt gctgtgctgc
901  gcttccacct tcttcaactgt caccacgtac ttggtagaca tgcagcgctt ccgctacca
961  gagcggccta tcatttttct gtcgggctgc tacaccatgg tgcgggtggc ctacatcgcg
1021  ggcttcgtgc tccaggagcg cgtgggtgtg aacgagcgct tctccgagga cggttaccgc
1081  acggtgggtg agggcaccaa gaaggagggc tgcaccatcc tcttcatgat gctctacttc
1141  ttcagcatgg ccagctccat ctgggtgggtc atcctgtcgc tcacctggtt cctggcagcc
1201  ggcataaagt ggggccacga ggccatcgag gccaaacttc agtacttcca cctggcgcgc
1261  tgggcccgtg cggccgtcaa gaccatcacc atcctggcca tgggccagat cgacggcgac
1321  ctgctgagcg gcgtgtgctt cgtaggcctc aacagcctgg acccgctgcg gggcttcgtg
1381  ctagegcgcg tcttcgtgta cctgttcate ggacgctcct tcctcctggc cggcttcgtg
1441  tcgctcttcc gcatccgcac catcatgaag cacgaaggca ccaagaccga aaagctggag
1501  cggctcatgg tgcgcatcgg cgtcttctcc gtgctctaca cagtgcgcgc caccatcgtc
1561  atcgcttgct acttctacga gcaggccttc cgcgagcact gggagcgctc gtgggtgagc
1621  cagcactgca agagcctggc catcccgtgc ccggcgact acacgcgcgc catgtcgccc
1681  gacttcacgg tctacatgat caaatacctc atgacgctca tcgtgggcat cacgtcgggc
1741  ttctggatct ggtcgggcaa gacgctgcac tcgtggagga agttctacac tcgcctcacc
1801  aacagccgac acggtgagac caccgtgtga gggacgcccc caggccggaa ccgcgcggcg
1861  ctttcctccg cccgggggtg ggcccctaca gactccgtat tttatttttt taaataaaaa
1921  acgacgaaa ccatttcact tttaggttgc tttttaaaag agaactctct gcccaacacc
1981  ccc (SEQ ID NO:95)

```

FIGURE 51A

85/115

Frizzled 2 (NM_001466)

MRPRSALPRLLLPLLLLPAAGPAQFHGEKGISIPDHGFCQPISI
PLCTDIAYNQTIMPNLLGHTNQEDAGLEVHQFYPLVKVQCSPELRFFLCSMYAPVCTV
LEQAIPPCRSICERARQGCEALMNKFGFQWPERLRCEHFPRHGAEQICVGQNHSEDGA
PALLTTAPPPGLQPGAGGTPGGPGGGGAPPRYATLEHPPFHCPRVLKVPSYLSYKFLGE
RDCAAPCEPARPDGSMFFSQEETRFARLWILTWSVLCCASTFFTVTTYLVDMQRFRYP
ERP IIFLSGCYTMVSVAYIAGFVLQERVVCNERFSEDGYRTVVQGTKKEGCTILFMML
YFFSMASSIWWVILSLTWFLAAGMKWGHEAIEANSQYFHAAWAVPAVKTTITILAMGQ
IDGDLLSGVCFVGLNSLDPLRGFVLAPLFVYLFIGTSFLLAGFVSLFRIRTIMKHDGT
KTEKLERLMVRIGVFSVLYTVPATIVIACYFYEQAFREHWERSWVSQHCKSLAIPCPA
HYTPRMSPDFTVYMIKYLMTLIVGITSGFWIWSGKTLHSWRKFYTRLTNSRHGETTV (SEQ ID NO:96)

FIGURE 51B

86/115

ISLR (NM_005545)

```

1  aagcagttgt tttgctggaa ggagggagtg cgcgggctgc cccgggctcc tccctgccgc
61  ctctctctcag tggatggttc caggcaccct gtctggggca gggaggggcac aggcctgcac
121  atcgaagggtg ggggtgggacc aggcctgcccc tcgccccagc atccaagtcc tcccttgggc
181  gcccggtggcc ctgcagactc tcagggtctaa ggtcctctgt tgcttttttg ttccacctta
241  gaagagggtc cgcttgacta agagtagctt gaaggaggca ccatgcagga gctgcatctg
301  ctctggtggg cgcttctcct gggcctggct caggcctgcc ctgagccctg cgactgtggg
361  gaaaagtatg gcttccagat cgccgactgt gcctaccgcg acctagaatc cgtgccgcct
421  ggcttcccg gccaatgtgac tacactgagc ctgtcagcca accggctgcc aggcctgccg
481  gaggggtgct tcaggggagg gcccctgctg cagtcgctgt ggctggcaca caatgagatc
541  cgcacggtgg ccgccggagc cctggcctct ctgagccatc tcaagagcct ggacctcagc
601  cacaatctca tctctgactt tgccctggagc gacctgcaca acctcagtgc cctccaattg
661  ctcaagatgg acagcaacga gctgaccttc atcccccgcg acgccttccg cagcctccgt
721  gctctgcgct cgctgcaact caaccacaac cgcttgacac cattggccga gggcaccttc
781  accccgctca ccgcgctgtc ccacctgcag atcaacgaga accccttcga ctgcacctgc
841  ggcacgtgtg ggctcaagac atggggccctg accacggcgg tgtccatccc ggagcaggac
901  aacatcgctt gcacctcacc ccatgtgctc aagggtacgc cgtgagccg cctgccgcca
961  ctgccatgct cggcgccctc agtgacgctc agctaccaac ccagccagga tggtgccgag
1021  ctgcgccctg gttttgtgct ggcactgcac tgtgatgtgg acgggcagcc gggccctcag
1081  cttcactggc acatccagat acccagtggc attgtggaga tcaccagccc caacgtgggc
1141  actgatgggc gtgccctgcc tggcaccctt gtggccagct cccagccgcg cttccaggcc
1201  tttgccaatg gcagcctgct tatccccgac tttggcaagc tggaggaagg cacctacagc
1261  tgcctggcca ccaatgagct gggcagtgct gagagctcag tggacgtggc actggccacg
1321  cccggtgagg gtggtgagga cacctgggg cgcagggttc atggcaaagc ggttgaggga
1381  aagggtgct atacggttga caacgaggtg cagccatcag ggcgggagga caatgtggtc
1441  atcatctacc tcagccgtgc tgggaacctt gaggtgcag tcgcagaagg ggtccctggg
1501  cagctgcccc caggcctgct cctgctgggc caaagcctcc tctcttctt cttcctcacc
1561  tccttctagc cccacccagg gcttccctaa ctctccctt tgccctacc aatgcccctt
1621  taagtgtgct aggggtctgg ggttggcaac tctgaggcc tgcattgggtg acttcacatt
1681  ttctacctc tccttctaat ctcttctaga gcacctgcta tccccactt ctagacctgc
1741  tccaaactag tgactaggat agaatttgat ccctaactc actgtctgcg gtgctcattg
1801  ctgctaacag cattgcctgt gctctcctct caggggcagc atgctaacgg ggcgacgtcc
1861  taatccaact gggagaagcc tcagtgggtg aattccaggc actgtgactg tcaagctggc
1921  aaggggcagg attgggggaa tggagctggg gcttagctgg gaggtggtct gaagcagaca
1981  gggaatggga gaggaggatg ggaagtagac agtggctggg atggctctga ggctccctgg
2041  ggcctgctca agctcctcct gctccttgct gtttctgat gatttggggg cttgggagtc
2101  cctttgtcct catctgagac tgaaatgtgg ggatccagga tggcttctt cctcttacc
2161  ttctccctc agcctgcaac ctctatcctg gaacctgtcc tcccttctc cccaactatg
2221  catctgttgt ctgctcctct gcaaaggcca gccagcttgg gagcagcaga gaaataaaca
2281  gcatttctga tgcc (SEQ ID NO:97)

```

FIGURE 52A

87/115

ISLR (NM_005545)

MQELHLLWWALLLGLAQACPEPCDCGEEKYGFQIADCAYRDLESV
PPGFANVTTLSSLNRLPGLPEGAFREVPLLQSLWLAHNEIRTVAAGALASLSHLKS
LDLSHNLISDFAWSDLHNLNLSALQLLKMSNELTFIPRDAFRSLRALRSIQLNHNRLHT
LAEGTFTPLTALSHLQINENPFDCGIVWLKTWALTAVSIPEQDNIACTSPHVLKG
TPLSRLPPLPCSAFVQLSYQPSQDGAELRPGFVLALHCDVDGQPAPQLHWHIQIPSG
IVEITSPNVGTDGRALPGTFVASSQPRFQAFANGSLLI PDFGKLEEGTYSCLATNELG
SAESSVDVALATPGEGGEDTLGRRFHGKAVEGKGCYTVDNVQPSGPEDNVVVIYLSR
AGNPEAAVAEGVPGQLPPGLLLLGQSLLLFFFLTSE (SEQ ID NO:98)

FIGURE 52B

88/115

FLJ23399 (NM_022763)

```

1  tgacccgggtc cgtgtggggcc agcgggaagg aagccagttg aggggaagttc tccatgaatg
61  tacgtcacaa tgatgatgac cgaccaaadc cctctggaac tgccaccatt gctgaacgga
121 gaggtagcca tgatgccccca cttggtgaat ggagatgcag ctcagcaggt tattctcggt
181 caagttaatc caggtgagac tttcacataa agagcagagg atggaacact tcagtgcatt
241 caaggacctg ctgaagttcc catgatgtca cccaatggat ccattcctcc cattcatgtg
301 cctccagggt atattctcaca ggtgattgaa gatagtactg gagtccgccc ggtggtgggtc
361 acaccccagt ctccctgagtg ttatcccca agctaccctc cagccatgtc tccaacccat
421 catctccctc cctatctgac tcaccatcca cattttattc ataactcaca cacggcttac
481 taccacactg ttaccggacc tggagatatg ccgcctcagt tttttcccca gcatcatctt
541 cccacacaaa tatatggtga gcaagaaatt ataccatttt atggaatgtc aagctacatc
601 acccgagaag accagtacag caagcctccg cacaaaaaac tgaaagaccg ccagatcgat
661 cgccagaacc gactcaacag acctccttct gctatctaca aaagcagctg cacaacagta
721 tacaatggct atgggaaggg ccatagtggg ggaagtggcg gaggcggcag cggtagtggg
781 cccggaatta agaaaacaga gcgacgagca agaagcagcc caaagtcgaa tgattcagac
841 ttgcaagaat atgagttgga agtaagagg gtgcaagaca ttcttccggg aatagagaaa
901 ccacaggttt ctaatatcca ggcaagagca gttgtgtgtg cctgggctcc ccctgttggg
961 ctttccctgtg gacccacacag tggctcttcc ttcccttaca gttacgaggt ggccttatca
1021 gacaaaggac gagatggaaa atacaagata atttacagtg gagaagaatt agaatgtaac
1081 ctgaaagatc ttagaccagc aacagattat catgtgaggg tgtatgccat gtacaattcc
1141 gtaaagggat cctgctccga gcctgttagc ttcaccaccc acagctgtgc acccgagtgt
1201 cttttccccc ctaagctggc acataggagc aaaagttcac taacctgca gtggaaggca
1261 ccaattgaca acggttcaaa aatcaccaac taccttttag agtgggatga gggaaaaaga
1321 aatagtgggt tcagacagtg cttcttcggg agccagaagc actgcaagtt gacaaagctt
1381 tgtccggcaa tggggtacac attcaggctg gccgctcgaa acgacattgg accagtggt
1441 tatagccaag aggtggtgtg ctacacatta ggaaatatcc ctcatagtc ttctgcacca
1501 aggtctggtc gagctggcat cacatgggtc acgttgcagt ggagtaagcc agaaggctgt
1561 tcacccgagg aagtgatcac ctacaccttg gaaattcagg aggatgaaaa tgataacctt
1621 ttccacccaa aatacactgg agaggattta acctgtactg tgaaaaatct caaaagaagc
1681 acacagtata cattcaggct gactgcttct aatacgggaag gaaaaagctg tccaagcgaa
1741 gttcttgttt gtacgacgag tctgacagg cctggacctc ctaccagacc gcttgtcaaa
1801 ggcccagtta catctcatgg ctttagtgtc aaatgggatc cccctaagga caatggtggg
1861 tcagaaatcc tcaagtactt gctagagatt actgatggaa attctgaagc gaatcagtg
1921 gaagtggcct acagtgggtc ggctaccgaa tacaccttca cccacttgaa accaggcact
1981 ttgtacaaac tccgagcatg ctgcatcagt accggcggac acagccagtg ttctgaaagt
2041 ctccctgttc gcacactaag cattgcacca ggtcaatgtc gaccaccgag ggttttgggt
2101 agaccaaagc acaaagaagt ccacttagag tgggatgttc ctgcatcgga aagtggctgt
2161 gaggtctcag agtacagcgt ggagatgacg gagcccgaag acgtagcctc ggaagtgtac
2221 catggcccag agctggagtg caccgtcggc aacctgcttc ctggaaccgt gtatcgcttc
2281 cgggtgaggg ctctgaatga tggaggggat ggtccctatt ctgatgtctc agaaattacc
2341 actgctgcag ggccctcctg acaatgcaaa gcaccttgta tttcttgtag acctgatgga
2401 tgtgtcttag tgggttggga gagtctgat agttctggtg ctgacatctc agagtacagg
2461 ttggaatggg gagaagatga agaactccta gaactcattt atcatgggac agacaccgt
2521 tttgaaataa gagacctgtt gcctgctgca cagtattgot gtagactaca ggccctcaat
2581 caagcagggg cagggccgta cagtgaactt gtcctttgcc agacgccagc gtctgcccct
2641 gaccccgctc ccactctctg tgtcctggag gaggagcccc ttgatgccta ccctgattca
2701 ccttctgcgt gccttgtagt gaactgggaa gagccgtgca ataaccggatc tgaaatcctt
2761 gcttacacca ttgatctagg agacactagc attaccgtgg gcaacaccac catgcatgtt
2821 atgaaagatc tccttccaga aaccacctac cggatcagaa ttcaggetat aaatgaaatt
2881 ggagctggac catttagtca gttcattaaa gcaaaaactc ggccattacc acccttgcc
2941 cctaggctag aatgtgctgc tgctggctct cagagcctga agctaaaatg gggagacagt
3001 aactccaaga cacatgctgc tgaggacatt gtgtacacac tacagatgga ggcagaaac
3061 aagaggttta tttcaatcta cagaggaccc agccacacct acaaggcca gagactgacg
3121 gaattcacat gctactcctt cagaatccag gcagcaagcg aggtggaga agggcccttc
3181 tcagaaacct atacctcag cacaaccaa agtgtcccc ccaccatcaa agcacctcga
3241 gtaacacagt tagaaggaaa tcatgtgaa attttatggg agacggtacc atcaatgaaa
3301 ggtgaccctg ttaactacat tctgcaggta ttggttggaa gagaatctga gtacaaacag

```

FIGURE 53A

89/115

```

3361 gtgtacaagg gagaagaagc cacattccaa atctcaggcc tccagaccaa cacagactac
3421 aggttccgcg tatgtgcgtg tctgcgtgt ttagacacct ctcaggagct aagcggagcc
3481 ttcagccctt ctgcggcttt tgtattacaa cgaagtgagg tcatgcttac agggggacatg
3541 gggagcttag atgatcccaa aatgaagagc atgatgccta ctgatgaaca gtttgcagcc
3601 atcattgtgc ttggctttgc aactttgtcc attttatttg cctttatatt acagtacttc
3661 ttaatgaagt aaacccaaca aaactagagg tatgaattaa tgctacacat ttaatacac
3721 acatttattc agatactccc ctttttaaag gtgttttggt gcaccttttt gaaatgcaaa
3781 tttacagatt tagctagaaa aaaaatgtca gtgttttggt gcaccttttt gaaatgcaaa
3841 actaggaaaa ggttaaactg gatttttttt tttaaaaaaa agaaaaaaa agaagaaaag
3901 tataccagat accaaaagct agctttctta tgttttcctt taaattttca gatttacctt
3961 cattctgttt tcaactgatgt cttttgcaag cctttgattt tttttttttt gttacagttt
4021 agtaatttat attcaccagt cacttcatat gtcttgaaca tctgtatctg taaacatgaa
4081 tcaccgtgtg tgtacttaca gggctaggat ttcagtgttg tcagagtatt accacacagc
4141 aacagcaaca tacagaagat atgttcactc agataagact gccctaaaca accattttgt
4201 cactcagtta tttaaactgtg tttagctcat ttaaatacaa atgtgtactt taacttaaaa
4261 tgttttaata atctgtattt cttataaatt taacactatg agctgcctgt ataagaaatc
4321 aagtaaccag aatgcaccta taaattatgg agcattgtag attttaccac atcaattcat
4381 agcagtaact ttaagagggc attgtgcaat agttagtgtt tttcttggtc agctatttta
4441 aaggctgctt taacttggtt gtttgccttt gtatataact acttctaatac taactactag
4501 agttattata ttctgttatg tttgaccaga attatatgac aagaactggg gacagtttag
4561 tgcctctgcc cattgtccat gatttacact aattgtgagc agtcttctta tgtgtcagct
4621 cattattttt gaaacatttg cctttaggct gttctttgag gtatcaatga agtgattgaa
4681 tttcaatacc ttaattcagt gcacataata ctaatgtaac agcagatgaa aattgataaa
4741 acccaaaaaga gagtcatcta aattttagt tccctatttct gtgggtttgc tggccatgg
4801 ttggagaggg aatgggtgtt gatggtaaac ctaacttgct gcgtgacctg aacacgtcac tttacctctc
4861 ttctctccct ggatctgtca catgaaaaat aaaataaaat aaaacgggga ttctaattgtt
4921 tgtgcctcag ttttcccatg tgaccaacag gtgctattgg agtgcaaagt gttactctta
4981 tgtaagtgtt ttgagatctt agataatcaa ttttaaccca aagtcatggg attattttata
5041 cgtgtttatt ttgagtcagt tacctcaggg acatttaaga gttggagggtg caaatatatt
5101 tgaagtccat aatgttcgag taccctcaggg ccttgcctct gtttttgtat atttttgtca
5161 ccaaaagggg gcaacagaca gctattttgt gcttgatctt tattgtctaa gatgcagtat
5221 cttgggtttt cttgatcata tcccatacca agtcatggg gaaacaaaca ttattttgtt
5281 cctgtactag cttataatat ttctgcatac agtactttaa atgccaatca cagtgcatac
5341 tttgggtttt gtataactata ttaagtgtac ttatgtacta attttccctt gtagcatgtt
5401 tttattttatt gtataactata tttgtaatgt taggtcagtc ttgttccctg gcaacatctg
5461 atatttttgt gttttatact tttgtaatgt taggtcagtc ttgttccctg gcaacatctg
5521 tagtattatt aatcttctga ctttttctta tgtttttaaa aagataagag catctagtgc
5581 attaaatgcc aaaaaaaaaa tacattatca gtgattgaaa cgtttacatg tacccaaaaa
5641 ccataatcat ctcttggaag aaaatgctga gatcaatgaa ttattctgtg tgcctatatt
5701 gacgtagtga gtactagaga gttctgtatt ttattattga ctataataat tagtttaatt
5761 agctttgcaa actgatggca tcaaggtaaa tataattttg ccaagttctt ggccttccaa
5821 aactcaccct cttattttaa tgtgtgctat gaccactatg tctgcatttt ttaagcaatg
5881 ctaaaaaatt ccatgcaggt gttttgggga gaggtatttt ttaagcaatg aaaattcaac
5941 tgagtacaaa gcccctctt ggggggttgg ggaagtctct tttttgaaac acttcagaac
6001 tgctgctata aagaaattct ctaattgggt gaattttttt ttttaagtaa tagtacttta
6061 ggccaaaatt tatatgaata tttgatcttc ttgagatttt catactatca ttttaaccac
6121 aggaagctga agtgtgtgaa gtacaaagct gacagcactt tattttattg ctctccatta
6181 tttgggtatt atttatattc ttcagtcaga aaattattac tctctatggc actgtttttt
6241 atcacaaaata tgtatatgtg atattgatata ataactatat atattgccat cacacacgaa
6301 caataaaaata aagtgttcta ttaacctgat ctctttgccc ttttgctatg tgaggagtga
6361 atgagtggcc ttctgatgct ctgactcttc tctgtatgtc aaactcatcc ctggcacaag
6421 aaattccagt catgtgaagc aaactgccct ttgtcctcaa agaaattgtt gaaaaagaaa
6481 acttttttaa gagatttttt gcatattctc tgccttggtc ttatcaactt gaaatgttgg
6541 catttttctaa ccttgttttg ttggctacaa taattcagta ttcatgtcaa aattgagaag
6601 tgccctaatt gaatgtgttt gaatgttatc cttgcacaat tctttaaatt gaaagataaa
6661 atgttttacc tcaactgttg acatacatct caagcttttc aactctagga gaaaaagaaa
6721 atcatgtttt cctgtattgt aaattttaga ctatttcata tacattgtat taaaactgcc
6781 atatcaattt taatgtatag attttgcaaa tattatgcta tatgtaatac ctaactgtat
6841 ctgtagtgta tatgtaatat atttatgccc aataaatgtt ttaattcttt ctga (SEQ ID

```

NO: 99)

FIGURE 53B

90/115

FLJ23399 (NM_022763)

MYVTMMMTDQIPLELPPLLNGEVAMMPHLVNGDAAQQVILVQVN
PGETFTTIRAEDGTLQCIQGPAEVPMMSPNGSIPPIHVPPGYISQVIEDSTGVRRVVVT
PQSPECYPPSYPSAMSPTHHLPPYLTHHPHFHNSHTAYYPPVTGPGDMPPQFFPQHH
LPHTIYGEQEIIIPFYGMSSYITREDQYSKPPHKKLKDRQIDRQNRLNRPPSAIYKSSC
TTVYNGYGKGHSGSGGGSGSGPGIKKTERRARSSPKSNDSDLQEYELEVKRVQDIL
SGIEKPQVSNIQARAVVLSWAPPVGLSCGPHSGLSFPYSYEVALSDKGRDGKYKIIYS
GEELECNLKDRLRPATDYHVRVYAMYNVKGSCSEPVSFTHSCAPECPFPKLAHRSK
SSLTLQWKAPIDNGSKITNYLLEWDEGKRNSGFRQCFFGSQKHCKLTKLCPAMGYTFR
LAARNDIGTSGYSQEVVCYTLGNIPQMPSAPRLVRAGITWVTLQWSKPEGCSPEEVIT
YTLEIQEDENDNLFHPKYTGEDLTCTVKNLKRSTQYTFRLTASNTEGKSCPSEVLVCT
TSPDRPGPPTRPLVKGPVTSHGFSVKWDPPKDNNGSEILKYLLEITDGNSEANQWEVA
YSGSATEYTFTHLKPGLTYKLRACCISTGGHSQCSESLPVRTL SIAPGQCRPPRVLGR
PKHKEVHLEWDVPASESGCEVSEYSVEMTEPEDVASEVYHGPELECTVGNLLPGTVYR
FRVRALNDGGYGPYSADVSEITTAAGPPGQCKAPCISCTPDGCVLVGWESPDSSGADIS
EYRLEWGEDEESLELIYHGTDRFEIRDLLPAAQYCCRLQAFNQAGAGPYSELVLCQT
PASAPDPVSTLCVLEEEPLDAYPDSPSACLVLNWEPCNNGSEILAYTIDLGDTSITV
GNTTMHVMKDLLPETTYRIRIQAINIEIGAGPFSQFIKAKTRPLPPLPPRLECAAAGPQ
SLKLKWGDSNSKTHAAEDIVYTLQLEDNRKRFISIRGPSHTYKVQRLTEFTCYSFRI
QAASEAGEGPFSETYTFSTTKSVPTTIKAPRVTQLEGNSCEILWETVPSMKGDPVNYI
LQVLVGRESEYKQVYKGEEATFQISGLQTN TDYRFRVCACRRCLDTSQELSGAFSPSA
AFVLQRSEVMLTGDMGSLDDPKMKSMMP TDEQFAAIIVLGFATLSILFAFILQYFLMK (SEQ ID NO:100)

FIGURE 53C

91/115

TEM1 (NM_020404)

```

1  tcgcgatgct gctgogcctg ttgctggcct gggcggcgcg agggcccaca ctggggccagg
61  acccctgggc tgetgagccc cgtgccgcct ggggccccag cagctgctac gctctcttcc
121  cacggcgccg cacccttccg gaggcctggc gggcctgccc cgagctgggg ggcgacctgg
181  cactcctcctg gacccccgag gagggccagc gtgtggacag cctgggtggg gcgggcccag
241  ccagccggct gctgtggatc gggctgcagc ggcaggcccc gcaatgccag ctgcagcgcc
301  cactgcgcgg cttcacgtgg accacagggg accaggacac ggctttcacc aactggggccc
361  agccagcctc tggaggcccc tgcccggccc agcgctgtgt ggccctggag gcaagtggcg
421  agcaccgctg gctggagggc tcgtgcacgc tggctgtcga cggctacctg tgccagtttg
481  gcttcgaggg cgccctgccc gcgctgcaag atgaggcggg ccaggccggc ccagccgtgt
541  ataccacgcc cttccacctg gtctccacag agtttgagtg gctgcccttc ggctctgtgg
601  ccgctgtgca gtgccaggct ggcaggggag cctctctgct ctgcgtgaag cagcctgagg
661  gaggtgtggg ctggtcacgg gctgggcccc tgtgcctggg gactggctgc agccctgaca
721  acgggggctg cgaacacgaa tgtgtggagg aggtggatgg tcacgtgtcc tgccgctgca
781  ctgagggctt ccggctggca gcagacgggc gcagttgcga ggaccctgtg gccaggctc
841  cgtgcgagca gcagtgtgag cccgggtggc cacaaggcta cagctgccac tgtgcctgg
901  gtttccggcc agcggaggat gatccgcacc gctgtgtgga cacagatgag tgccagattg
961  ccgggtgtgt ccagcagatg tgtgtcaact acgttggtgg cttcgagtgt tattgtagcg
1021  agggacatga gctggaggct gatggcatca gctgcagccc tgcaggggccc atgggtgccc
1081  aggcctccca ggacctcgga gatgagttgc tggatgacgg ggaggatgag gaagatgaag
1141  acgaggcctg gaaggccttc aacgggtggc ggacggagat gcctgggatc ctgtggatgg
1201  agcctacgca gccgcctgac tttgccctgg cctatagacc gagcttccca gaggacagag
1261  agccacagat accctacccg gagcccacct ggccaccccc gctcagtgcc cccagggtcc
1321  cctaccactc ctcaagtgtc tccgtcacc ggccctgtgg ggtctctgcc acgcatccca
1381  cactgccttc tgcccaccag cctcctgtga tccctgccac acacccagct ttgtcccgtg
1441  accaccagat ccccgatgat gcagccaact atccagatct gccttctgcc taccaaccgg
1501  gtattctctc tgtctctcat tcagcacagc ctcccgccca ccagccccct atgatctcaa
1561  ccaaatatcc ggagctcttc cctgccacc agtcccccat gtttcagacc acccggtcg
1621  ctggcaccca gaccaccact catttgcttg gaatccacc taaccatgcc cctctgggtc
1681  ccaccctcgg tgcccagcta cccctcaag cccagatgc ccttgtctc agaaccagg
1741  ccaccagct tccattatc ccaactgcc agccctctct gaccaccacc tccaggtccc
1801  ctgtgtctcc tgcccatcaa atctctgtgc ctgtgccac ccagcccgca gccctcccca
1861  ccctcctgcc ctctcagagc cccactaacc agacctcacc catcagccct acacatcccc
1921  attccaaagc cccccaatc ccaagggaag atggccccag tcccaagttg gccctgtggc
1981  tgccctcacc agctcccaca gcagcccaa cagccctggg ggaggctggg cttgccgagc
2041  acagccagag ggatgaccgg tggctgctgg tggcactcct ggtgccaacg tgtgtctttt
2101  tgggtggctc gcttgcaact ggcatcgtgt actgcacccg ctgtggcccc catgcacca
2161  acaagcgcac cactgactgc tatcgtggg tcatccatgc tgggagcaag agcccaacag
2221  aacccatgcc ccccaggggc agcctcacag ggggtgcagac ctgcagaacc agcgtgtgat
2281  ggggtgcaga cccctctcat ggagtatgg gcgctggaca catggccggg gctgcaccag
2341  ggaccatgg gggctgccc gctggacaga tggcttctg ctcccaggc ccagccaggg
2401  tcctctctca accactagac ttggctctca ggaactctgc ttcctggccc agcgtcgtg
2461  accaaggata caccaaagcc cttaagacct cagggggcgg gtgctggggg cttctccaat
2521  aaatgggggtg tcaaccttaa aaaaaaaaaa aaaaaaaaaa aaaaa (SEQ ID NO:101)

```

FIGURE 54A

92/115

TEM1 (NM_020404)

MLLRLLLLAWAAAGPTLGQDPWAAEFRAACGPSSCYALFPRRRTF
LEAWRACRELGGDLATPRTPEEAQRVDSL VGAGPASRLWIGLQRQARQCQLQRPLRG
FTWTTGDQDTAFTNWAQPASGGPCPAQRCVALEASGEHRWLEGSCTLAVDGYLCQFGF
EGACPALQDEAGQAGPAVYTTTFHLVSTEFEWLPFGSVAAVQCQAGRGASLLCVKQPE
GGVGWSRAGPLCLGTGCS PDNGGCEHECVEEVDGHVSCRCTEGFRLAADGRSCEDPCA
QAPCEQQCEPGGPQGYSCHCRLGFRPAEDDPHRCVDTDECQIAGVCQQMCVNYVGGFE
CYCSEGELEADGISCPAGAMGAQASQDLGDELLDDGEDEEDEDEAWKAFNGGWTEM
PGILWMEFTQPPDFALAYRPSFPEDREPQIPYPEPTWPPPLSAPRVPHYSSVLSVTRP
VVVSATHPTLPSAHQPPVIPATHPALSRDHQIPVIAANYPDLP SAYQPGILSVSHSAQ
PPAHQPPMISTKYPELFPAHQSPMFPDTRVAGTQTTHLPGIPPNHAPLVTTLGAQLP
PQAPDALVLR TQATQLPIIPTAQPSLT TTSRSPVSPAHQISVPAATQPAALPTLLPSQ
SPTNQTSPI SPTHPHSKAPQIPREDGSPSKLALWLPS PAPTAAPTALGEAGLAEHSQR
DDRWLLVALLVPTCVFLVLLALGIVYCTRCGPHAPNKRITDCYRWVIHAGSKSPTEP
MPPRGSLTGVQTCRTSV (SEQ ID NO:102)

FIGURE 54B

93/115

Tie2 ligand2 (NM_001147)

```

1  tgggttggtg tttatctcct cccagccttg agggagggaa caacactgta ggatctgggg
61 agagaggaac aaaggaccgt gaaagctgct ctgtaaaagc tgacacagcc ctcccaagt
121 agcaggactg ttcttccac tgcaatctga cagtttactg catgcctgga gagaacacag
181 cagtaaaaac caggtttgct actggaaaaa gaggaagag aagactttca ttgacggacc
241 cagccatggc agcgtagcag ccctgcgttt cagacggcag cagctcggga ctctggacgt
301 gtgtttgccc tcaagtttgc taagctgctg gtttattact gaagaaagaa tgtggcagat
361 tgttttcttt actctgagct gtgatcttgt cttggccgca gcctataaca actttcggaa
421 gagcatggac agcataggaa agaagcaata tcaggtccag catgggtcct gcagctacac
481 tttcctcctg ccagagatgg acaactgccg ctcttcctcc agcccctacg tgtccaatgc
541 tgtgcagagg gacgcgcgcg tcgaatacga tgactcgggt cagaggtgcg aagtgcctgga
601 gaacatcatg gaaaacaaca ctcagtggct aatgaagctt gagaattata tccaggacaa
661 catgaagaaa gaaatggtag agatacagca gaatgcagta cagaaccaga cggctgtgat
721 gatagaaata gggacaaacc tgttgaaaca aacagctgag caaacgcgga agttaactga
781 tgtggaagcc caagtattaa atcagaccac gagacttgaa cttcagctct tggaaactc
841 cctctcgaca aacaaattgg aaaaacagat tttggaccag accagtgaat taaacaaatt
901 gcaagataag aacagtttcc tagaaaagaa ggtgctagct atggaagaca agcacatcat
961 ccaactacag tcaataaaaag aagagaaaga tcagctacag gtgttagtat ccaagcaaaa
1021 ttccatcatt gaagaactag aaaaaaaaaa agtgactgcc acggtgaata attcagttct
1081 tcaaaaggag caacatgac tcattggagac agttaataac ttactgacta tgatgtccac
1141 atcaaactca gctaaggacc ccactgttgc taaagaagaa caaatcagct tcagagactg
1201 tgctgaagta ttcaaatcag gacacaccac aaatggcatc tacacgttaa cattccctaa
1261 ttctacagaa gagatcaagg cctactgtga catggaagct ggaggaggcg ggtggacaat
1321 tattcagcga cgtgaggatg gcagcgttga ttttcagagg acttggaag aatataaagt
1381 gggatttggt aacccttcag gagaatattg gctgggaaat gagtttgttt cgcaactgac
1441 taatcagcaa cgctatgtgc ttaaaataca ccttaaagac tgggaaggga atgaggctta
1501 ctcatgtgat gaacatttct atctctcaag tgaagaactc aattatagga ttcaccttaa
1561 aggacttaca gggacagccg gcaaaataag cagcatcagc caaccaggaa atgatttttag
1621 caaaaaggat ggagacaacg acaaatgtat ttgcaaatgt tcacaaatgc taacaggagg
1681 ctggtggttt gatgcatgtg gtccttccaa cttgaacgga atgtactatc cacagaggga
1741 gaacacaaat aagttcaacg gcattaaatg gtactactgg aaaggctcag gctattcgct
1801 caaggccaca accatgatga tccgaccagc agatttctaa acatcccagt ccacctgagg
1861 aactgtctcg aactattttc aaagacttaa gccagtgca ctgaaagtca cggctgcgca
1921 ctgtgtcctc ttccaccaca gagggcgtgt gctcgggtgt gacgggaccc acatgctcca
1981 gattagagcc tgtaaacttt atcacttaaa cttgcatcac ttaacggacc aaagcaagac
2041 cctaaacatc cataattgtg attagacaga acacctatgc aaagatgaac ccgaggctga
2101 gaatcagact gacagtttac agacgctgct gtcacaacca agaattgtat gtgcaagttt
2161 atcagtaaat aactggaaaa cagaacactt atgttatata atacagatca tcttggaact
2221 gcattcttct gagcactgtt tatacactgt gtaaatacc atatgtcct (SEQ ID

```

NO:103)

FIGURE 55A

94/115

Tie2 ligand2 (NM_001147)

MWQIVFFTLSCDLVLAAAYNNFRKSMDSIGKKQYQVQHGSCT
FLLPEMDNCRSSSSPYVSNVQRDAPLEYDDSVQRLQVLENIMENNTQWLMKLENYIQ
DNMKKEMVEIQQNAVQNQTAVMIEIGTNLLNQTAEQTRKLTDEAQVLNQTTRELEQL
LEHSLSTNKLEKQILDQTSEINKLQDKNSFLEKKVLAMEDKHIIQLQSIKEEKDQLQV
LVSKQNSIIIEELEKKIVTATVNNSVLQKQQHDLMETVNNLLTMMSTSNSAKDPTVAKE
EQISFRDCAEVFKSGHTTNGIYTTLTFPNSTEEIKAYCDMEAGGGGTIIQRREDGSVD
FQRTWKEYKVGFQNPGEYWLGNFVSQLTNQQRVYLKIHLKDWEAGNEAYSLEYEHFYL
SSEELNYRIHLKGLTG TAGKISSISQPGNDFSTKGDNDKCICKCSQMLTGGWWFDAC
GPSNLNGMYYPQRQNTNKFNGIKWYYWKGSGYSLKATMMIRPADF (SEQ ID NO:104)

FIGURE 55B

95/115

VEGFC (NM_005429)

```

1  cggggaaggg gagggaggag ggggacgagg gctctggcgg gtttggaggg gctgaacatc
61  gcggggtgtt ctggtgtccc ccgccccgcc tctccaaaaa gctacaccga cgcggaccgc
121  ggcggcgtcc tccctcgccc tcgcttcacc tcgcgggctc cgaatgcggg gagctcggat
181  gtccggtttc ctgtgaggct tttacctgac acccgccgcc tttccccggc actggctggg
241  agggcgccct gcaaagttag gaacgcggag ccccggaacc gctccccgcg cctccggctc
301  gccagggggg ggtcgccggg aggagcccgg gggagaggga ccaggagggg ccgcggcct
361  cgcaggggcg cccgcgcccc caccctgcc cccgccagcg gaccggtccc ccaccccg
421  tccttccacc atgcacttgc tgggcttctt ctctgtggcg tgttctctgc tcgccgtgc
481  gctgctcccg ggtcctcgcg aggcgcccg cgcgcgcgc gccttcgagt ccggactcga
541  cctctcggac gcggagcccg acgcgggcga ggccacggct tatgcaagca aagatctgga
601  ggagcagtta cggctctgtg ccagtgtaga tgaactcatg actgtactct acccagaata
661  ttggaaaatg tacaagtgtc agctaaggaa aggaggctgg caacataaca gagaacaggc
721  caacctcaac tcaaggacag aagagactat aaaatttgct gcagcacatt ataatacaga
781  gatcttgaaa agtattgata atgagtggag aaagactcaa tgcattgccac gggagggtgtg
841  tatagatgtg gggaaggagt ttggagtgcg gacaaacacc ttctttaaac ctccatgtgt
901  gtccgtctac agatgtgggg gttgctgcaa tagtgagggg ctgcagtgca tgaacaccag
961  cagcagctac ctcaagcaaga cgttatttga aattacagtg cctctctctc aaggcccca
1021 accagtaaca atcagttttg ccaatcacac ttcttgccga tgcattgtct aactggatgt
1081 ttacagacaa gttcattcca ttattagacg ttccctgcc a gcaacactac cacagtgtca
1141 ggcagcgaac aagacctgcc ccaccaatta catgtggaat aatcacatct gcagatgcct
1201 ggctcaggaa gattttatgt tttcctcgga tgctggagat gactcaacag atggattcca
1261 tgacatctgt ggaccaaaca aggagctgga tgaagagacc tgtcagtgtg tctgcagagc
1321 ggggcttcgg cctgccagct gtggaccca caaagaacta gacagaaact catgccagt
1381 tgtctgtaaa aacaaaactc tccccagcca atgtggggcc aaccgagaat ttgatgaaaa
1441 cacatgccag tgtgtatgta aaagaacctg cccagaaat caaccctaa atcctggaaa
1501 atgtgcctgt gaatgtacag aaagtccaca gaaatgcttg ttaaaaggaa agaagttcca
1561 ccaccaaaca tgcagctgtt acagacggcc atgtacgaac cgcagaagg cttgtgagcc
1621 aggattttca tatagtgaag aagtgtgtcg ttgtgtccct tcatattgga aaagaccaca
1681 aatgagctaa gattgtactg ttttccagtt catcgatttt ctattatgga aaactgtgtt
1741 gccacagtag aactgtctgt gaacagagag acccttgtgg gtccatgcta acaaagacaa
1801 aagtctgtct ttctgaacc atgtggataa ctttacagaa atggactgga gctcatctgc
1861 aaaaggcctc ttgtaaagac tggttttctg ccaatgacca aacagccaag attttctct
1921 tgtgatttct ttaaaagaat gactatataa tttatttcca ctaaaaatat tgtttctgca
1981 ttcattttta tagcaacaac aattggtaaa actcactgtg atcaatat tttatcatg
2041 caaaatatgt ttaaaataaa atgaaaattg tattat (SEQ ID NO:105)

```

FIGURE 56A

96/115

VEGFC (NM_005429)

MHLLGFFSVACSLLAALLPGPREAPAAAAAFESGLDLSDAEPD
AGEATAYASKDLEEQLRSVSSVDELMTVLYPEYWKMVKCQLRKGGWQHNREQANLSR
TEETIKFAAAHYNTEILKSIDNEWKKTQCMPEVCIDVGKEFGVATNTFFKPPCVSVY
RCGGCCNSEGLQCMNTSTSYLSKTLFEITVPLSQGPKPVTISFANHTSCRCMSKLDVY
RQVHSIIRRSLPATLPQQAANKTCPTNYMWNHICRCLAQEDFMFSSDAGDDSTDGF
HDICGPNKELDEETCQCVCRAGLRPASCGPHKELDRNSCQCVCKNKLFPSQCGANREF
DENTCQCVCCKRTCPRNQPLNPGKCAECTESPQKCLLKGGKFHHQTCSCYRRPCTNRQ
KACEPGFSYSEEVCRCVPSYWKRPM (SEQ ID NO:106)

FIGURE 56B

97/115

tPA (NM_000930)

```

1  atggccctgt  ccactgagca  tcctcccgcc  acacagaaac  ccgcccagcc  gggggccaccg
61  accccacccc  ctgcctggaa  acttaaggag  gccggagctg  tggggagctc  agagctgaga
121  tcctacagga  gtccagggct  ggagagaaaa  cctctgcgag  gaaagggaag  gagcaagccg
181  tgaatttaag  ggacgctgtg  aagcaatcat  ggatgcaatg  aagagagggc  tctgctgtgt
241  gctgctgctg  tgtggagcag  tcttcgtttc  gccagccag  gaaatccatg  cccgattcag
301  aagaggagcc  agatcttacc  aagtgatctg  cagagatgaa  aaaacgcaga  tgatatacca
361  gcaacatcag  tcatggctgc  gccctgtgct  cagaagcaac  cgggtggaat  attgctggtg
421  caacagtggc  agggcacagt  gccactcagt  gcctgtcaaa  agttgcagcg  agccaagggtg
481  tttcaacggg  ggcacctgcc  agcaggccct  gtacttctca  gatttctgtg  gccagtggcc
541  cgaaggatth  gctgggaagt  gctgtgaaat  agataccagg  gccacgtgct  acgaggacca
601  gggcatcagc  tacaggggca  cgtggagcac  agcggagagt  ggcgcagagt  gcaccaactg
661  gaacagcagc  gcgttggccc  agaagcccta  cagcgggcgg  aggccagacg  ccatcaggct
721  gggcctgggg  aaccacaact  actgcagaaa  cccagatcga  gactcaaagc  cctggtgcta
781  cgtctttaag  gcggggaagt  acagctcaga  gttctgcagc  acccctgcct  gctctgaggg
841  aaacagtgac  tgctactttg  ggaatgggtc  agcctaccgt  ggcacgcaca  gcctcaccga
901  gtcgggtgcc  tcctgcctcc  cgtggaattc  catgatcctg  ataggcaagg  tttacacagc
961  acagaacccc  agtgcaccag  cactgggcct  gggcaaacat  aattactgcc  ggaatcctga
1021  tggggatgcc  aagccctggg  gccacgtgct  gaagaaccgc  aggctgacgt  gggagtactg
1081  tgatgtgccc  tcctgctcca  cctgcggcct  gagacagtac  agccagcctc  agtttcgcat
1141  caaaggaggg  ctcttcgccc  acatcgccct  ccacccctgg  caggctgcca  tctttgccaa
1201  gcacaggagg  tcgcccggag  agcggttcct  gtgcgggggc  atactcatca  gctcctgctg
1261  gattctctct  gcgcccact  gcttcaggga  gaggtttccg  cccaccacc  tgacgggtgat
1321  cttgggcaga  acataccggg  tggccctgg  cgaggaggag  cagaaatttg  aagtcgaaaa
1381  atacattgtc  cataaggaat  tcgatgatga  cacttacgac  aatgacattg  cgctgctgca
1441  gctgaaatcg  gattcgctcc  gctgtgccc  ggagagcagc  gtggtccgca  ctgtgtgcct
1501  tcccccgcg  gacctgcagc  tgccggactg  gacggagtgt  gagctctccg  gctacggcaa
1561  gcatgaggcc  ttgtctcctt  tctattcgga  gcggctgaag  gaggtcatg  tcagactgta
1621  cccatccagc  cgctgcacat  cacaacatth  acttaacaga  acagtcaccg  acaacatgct
1681  gtgtgctgga  gacactcgga  gcggcggggc  ccaggcaaac  ttgcacgacg  cctgccaggg
1741  cgattcggga  ggccccctgg  tgtgtctgaa  cgatggccgc  atgactttgg  tgggcatcat
1801  cagctggggc  ctgggctgtg  gacagaagga  tgtcccgggt  gtgtacacca  aggttaccaa
1861  ctacctagac  tggattcgtg  acaacatgcg  accgtgacca  ggaacaccog  actcctcaaa
1921  agcaaagtag  atcccgccct  ttcttcttca  gaagacactg  caaaggcgca  gtgcttctct
1981  acagacttct  ccagacccac  cacaccgcag  aagcgggacg  agaccctaca  ggagaggggaa
2041  gagtgcattt  tcccagatac  ttccattttt  ggaagtthtc  aggaacttgg  ctgatttcag
2101  gatactctgt  cagatgggaa  gacatgaatg  cactactagc  tctccaggaa  tgccctctcc
2161  ctgggcagaa  agtggccatg  ccacccctgt  ttcagctaaa  gcccaacctc  ctgacctgtc
2221  accgtgagca  gctttggaaa  caggaccaca  aaaatgaaag  catgtotcaa  tagtaaaaga
2281  taacaagatc  tttcaggaaa  gacggattgc  attagaaata  gacagtatat  ttatagtcac
2341  aagagcccag  cagggccctc  aagttggggc  aggctggctg  gcccgctcat  ttcctcaaaa
2401  gcacccttga  cgtcaagtct  ccttcccctt  tcccactcc  ctggctctca  gaaggatttc
2461  cttttgtgta  cagtgtgtaa  agtgtaaatc  ctttttcttt  ataaacttta  gagtagcatg
2521  agagaattgt  atcatttgaa  caactaggct  tcagcatatt  tatagcaatc  catgttagtt
2581  tttactttct  gttgccacaa  cctgttttta  tactgtactt  aataaattca  gatatattht
2641  tcacagthtt  tcc (SEQ ID NO:107)

```

FIGURE 57A

98/115

tPA (NM_000930)

MDAMKRGLCCVLLLCGAVFVSPSQEIHARFRRGARSYQVICRDE
KTQMIYQQHQSWLRPVLRSNRVEYCWCNSGRAQCHSVPVKSCSEPRCFNGGTCQQALY
FSDFVCQCPEGFAGKCCEIDTRATCYEDQGISYRGTWSTAESGAECTNWNSSALAQKP
YSGRRPDAILRGLGNHNYCRNPDRDSKPWCYVFKAGKYSSEFCSTPACSEGNSDCYFG
NGSAYRGTHSLTESGASCLPWNSMILIGKVYTAQNPSAALGLGKHNYCRNPDGDAKP
WCHVLKNRRLTWEYCDVPSCSTCGLRQYSQPQFRIKGGLFADIASHPWQAAlFAKHRR
SPGERFLCGGILISSCWILSAAHCFQERFPPHLLTVILGRTYRVVPGEEEQKFEVEKY
IVHKEFDDDTYDNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTECELSGYG
KHEALSPFYSERLKEAHVRLYPSSRCTSQHLLNRTVTDNMLCAGDTRSGGPQANLHDA
CQGDSGGPLVCLNDGRMTLVGIISWGLGCGQKDVPGVYTKVTNYLDWIRDNM RP (SEQ ID NO:108)

FIGURE 57B

99/115

Thrombomodulin (NM_000361)

```

1  cttgcaatcc aggccttttct tggagagtggc tgtaacatgt atgaaaagaa agaaaggagg
61  accaagagat gaaagagggc tgcacgcgtg ggggcccagag tgggtgggccc ggacagtcgt
121 cttgttacag ggggtgctggc cttccctggc gcctgcccct gtccggccccg cccgagaacc
181 tccctgcgcc agggcagggg ttactcatcc cggcgaggtg atcccatgcg cgagggcggg
241 cgcaagggcg gccagagaa cagcaatcc gagtatgcg catcagccct tcccaccagg
301 cacttccttc cttttcccg aogtccaggg agggagggcc gggcacttat aaactcgagc
361 cctggccgat ccgcatgtca gaggtgcct cgcaggggct gcgcgcacgg caagaagtgt
421 ctgggctggg acggacagga gaggtgtcg ccatcggcgt cctgtgcccc tctgtccgg
481 cacggccctg tcgcagtgc cgcgtttcc ccggcgctg cacgcggcgc gcctgggtaa
541 catgcttggg gtccctggcc ttggcgcgct ggccctggcc ggccctgggg tccccgcacc
601 cgcagagccg cagccgggtg gcagccagtg cgtcgagcac gactgcttcg cgctctacc
661 gggccccgcg accttcctca atgccagtca gatctgcgac ggactgcggg gccacctaat
721 gacagtgcgc tcctcgggtg ctgccgatgt catttccttg ctactgaacg gcgacggcgg
781 cgttggcgcg cggcgccctc ggatcggcct gcagctgcca cccggctgcg gcgaccccaa
841 gcgcctcggg cccctgcgcg gcttcagtg gtttacggga gacaacaaca ccagctatag
901 caggtgggca cggctcgacc tcaatggggc tcccctctgc ggcccgctgt gcgtcgctgt
961 ctccgctgct gagggcactg tgcccagcga gccgatctgg gaggagcagc agtgcgaagt
1021 gaaggccgat ggcttcctct gcgagttcca cttcccagcc acctgcaggc cactggctgt
1081 ggagcccggc gccgcggctg ccgcgctctc gatcacctac ggcaccccg tgcggcccg
1141 cggagcggac ttccaggcgc tgccgggtgg cagctccgcc gcggtggctc ccctcggctt
1201 acagctaatt tgcaccgcgc cgcgcggagc ggtccagggg cactgggcca gggaggcgcc
1261 gggcgcttgg gactgcagcg tggagaacgg cggctgcgag cacgcgtgca atgcgatccc
1321 tggggctccc cgtgcccagt gccagccgg cgcgccttg caggcagacg ggcgctctg
1381 caccgatcc cgcagcagt cctgcaacga cctctgcgag cactctgcg tcccaacc
1441 gcaccagccg ggctcctact cgtgcatgtg cgagaccggc taccggctgg cggccgacca
1501 acaccggtgc gaggaagctg atgactgcat actggagccc agtccgtgtc cgcagcgtg
1561 tgtcaacaca cagggtggct tcgagtgcca ctgctaccct aactacgacc tgggtggcgg
1621 cgagtgtgtg gagcccgtgg acccgtgctt cagagccaac tgcgagtacc agtgccagcc
1681 cctgaaccaa actagctacc tctgctctg cgccgagggc ttcgcgcccc tccccacga
1741 gccgcacagg tgccagatgt tttgcaacca gactgcctgt ccagccgact gcgaccccaa
1801 caccagggct agctgtgagt gccctgaagg ctacatcctg gacgacgggt tcatctgcac
1861 ggacatcgac gactgcgaaa acggcggttc ctgctccggg gtgtgccaca acctccccg
1921 taccttcgag tgcattctgc ggcccgactc ggcccttgcc cgcccgactg gcaccgactg
1981 tgactccggc aagggtggac gtggcgacag cggctctggc gagccccgc ccagcccgac
2041 gcccggtccc accttgactc ctccggccgt ggggctcgtg cattcgggct tgctcatagg
2101 catctccatc gcgagcctgt gcctgggtgg ggcgcttttg gcgctcctct gccacctgcg
2161 caagaagcag ggcgcgcgca gggccaagat ggagtacaag tgcgcggccc cttccaagga
2221 ggtagtgtcg cagcacgtgc ggaccgagcg gacgccgag agactctgag cggcctcgt
2281 ccaggagcct ggctccgtcc aggagctgtg cctcctcacc cccagctttg ctaccaaagc
2341 acctagctg gcattacagc tggagaagac cctccccgca ccccccaagc tgttttctc
2401 tattccatgg ctaactggcg agggggtgat tagaggagg agaatgagc tcggcctctt
2461 ccgtgacgtc actggaccac tgggcaatga tggcaatttt gtaacgaaga cacagactgc
2521 gatttgtccc aggtcctcac taccggggcg aggaggggtg gcgttattgg tcggcagcct
2581 tctgggcaga ccttgacctc gtgggctagg gatgactaaa atatttattt tttttaagta
2641 tttaggtttt tgtttgtttc ctttgttctt acctgtatgt ctccagtatc cactttgcac
2701 agctctccgg tctctctctc tctacaaact cccacttgct atgtgacagg taaactatct
2761 tgggtgaattt ttttttctca gccctctcac atttatgaag caagccccac ttattcccca
2821 ttcttctctg ttttctctcc ccaggaactg ggccaactca cctgagtcac cctacctgtg
2881 cctgacccta cttcttttgc tcacttagct gtctgctcag acagaacccc tacatgaaac
2941 agaaacaaaa acactaaaaa taataatggc catttgcttt ttcaccagat ttgctaattt
3001 atcctgaaat ttcagattcc cagagcaaaa taatttttaa caaagggttg agatgtaaaa
3061 ggtattaaat tgatgttgcg ggactgtcat agaaattaca cccaaagagg tattttatctt
3121 tactttttaa cagtgagcct gaattttgtt gctgttttga tttgtactga aaaatggtaa
3181 ttgttgctaa tcttcttatg caatttcctt ttttgttatt attacttatt tttgacagtg
3241 ttgaaaatgt tcagaagggt gctctagatt gagagaagag acaaacacct cccaggagac
3301 agttcaagaa agcttcaaac tgcattgatt atgccaatga gcaattgact gtcactgttc

```

FIGURE 58A

100/115

```
3361 cttgtcactg gtagaccaa ataaaaccag ctctactggt cttgtggaat tgggagcttg
3421 ggaatggatc ctggaggatg cccaattagg gcctagcctt aatcaggtcc tcagagaatt
3481 tctaccatth cagagaggcc ttttggaatg tggccctga acaagaattg gaagctgccc
3541 tgcccatggg agctggttag aaatgcagaa tcctaggctc caccatcc agttcatgag
3601 aatctatatt taacaagatc tgcagggggg gtgtctgctc agtaatttga ggacaaccat
3661 tccagactgc ttccaatttt ctggaataca tgaaatatag atcagttata agtagcaggc
3721 caagtcaggc ccttattttc aagaaactga ggaattttct ttgtgtagct ttgctctttg
3781 gtagaaaagg ctaggtacac agctctagac actgccacac agggctgca aggtctttgg
3841 ttcagctaag ctaggaatga aatcctgctt cagtgtatgg aaataaatgt atcatagaaa
3901 tgtaactttt gtaagacaaa ggttttcctc ttctattttg taaactcaaa atatttgtac
3961 atagttatth atttattgga gataatctag aacacaggca aaatccttgc ttatgacatc
4021 acttgtacaa aataaacaaa taacaatgtg (SEQ ID NO:109)
```

FIGURE 58B

101/115

Thrombomodulin (NM_000361)

MLGVLVLGALALAGLGFPAPAEPPQPGGSQCVEHDCFALYPGPAT
FLNASQICDGLRGHLMTVRSSVAADVISLLLNQDGGVGRRLWIGLQLPPGCGDPKRL
GPLRGFQWVTGDNNTSYSRWARLDLNGAPLCGPLCVAVSAAEATVPSEPIWEEQQCEV
KADGFLCEFHFPAFCRPLAVEPGAAAAAVSITYGTPFAARGADFQALPVGSSAAVAPL
GLQLMCTAPPGAVQGHWAREAPGAWDCSVENGCEHACNAIPGAPRCQCPAGAALQAD
GRSCTASATQSCNDLCEHFCVPNPDQPGSYSCMCETGYRLAADQHRCEVDVDDCILEPS
PCPQRCVNTQGGFECHCYPNYDLVDGECVEPVDPCFRANCEYQCQPLNQTSYLCVCAE
GFAPIPHEPHRCQMFNQTACPADCDPNTQASCECPEGYILDDGFICTDIDECENGGF
CSGVCHNLPGTFCICGPDSEALRHIGTDCDSGKVDGGDSGSGEPPPSPTPGSTLTTP
AVGLVHSGLLIGISIASLCLVVALLLALLCHLRKKQGAARAKMEYKCAAPSKEVVLQHV
RTERTPQRL (SEQ ID NO:110)

FIGURE 58C

102/115

TF (NM_001993)

```

1 aagactgcga gctccccgca cccctcgca ctccctctgg ccggcccagg gcgccttcag
61 cccaacctcc ccagccccac gggcgccacg gaacccgctc gatctcgccg ccaactggta
121 gacatggaga cccctgcctg gccccgggtc ccgcgccccg agaccgccgt cgctcggacg
181 ctcttgctcg gctgggtctt cgcccagggtg gccggcgctt caggcactac aaatactgtg
241 gcagcatata atttaacttg gaaatcaact aatttcaaga caattttgga gtgggaaacc
301 aaacccgtca atcaagtcta cactgttcaa ataagcacta agtcaggaga ttggaaaagc
361 aaatgctttt acacaacaga cacagagtgt gacctcaccg acgagattgt gaaggatgtg
421 aagcagacgt acttggcacg ggtcttctcc taccgggcag ggaatgtgga ggcaccgggt
481 tctgctgggg agcctctgta tgagaactcc ccagagttca caccttacct ggagacaaac
541 ctcggaacgc caacaattca gagttttgaa cagggtggaa caaaagtga tgtgaccgta
601 gaagatgaac ggacttttagt cagaaggaac aacactttcc taagcctccg ggatgttttt
661 ggcaaggact taattttatac actttattat tggaaatctt caagttcagg aaagaaaaca
721 gccaaaacaa acactaatga gtttttgatt gatgtggata aaggagaaaa ctactgtttc
781 agtgttcaag cagtgattcc ctcccgaaca gttaaccgga agagtacaga cagcccggta
841 gagtgtatgg gccaggagaa aggggaattc agagaaatat tctacatcat tggagctgtg
901 gtattttgtg tcatcatcct tgtcatcatc ctggctatat ctctacacaa gtgtagaaag
961 gcaggagtgg ggcagagctg gaaggagaac tcccactga atgtttcata aaggaagcac
1021 tgttggagct actgcaaatg ctatattgca ctgtgaccga gaacttttaa gaggatagaa
1081 tacatggaaa cgcaaatgag tatttcggag catgaagacc ctggagttca aaaaactctt
1141 gatatgacct gttattacca ttagcattct ggttttgaca tcagcattag tcactttgaa
1201 atgtaacgaa tggtagtaca accaattcca agttttaatt tttaacacca tggcaccttt
1261 tgcacataac atgcttttaga ttatatattc cgcacttaag gattaaccag gtcgtccaag
1321 caaaaacaaa tgggaaaatg tcttaaaaaa tcttgggtgg acttttgaaa agcttttttt
1381 tttttttttt tttgagacgg agtcttgctc tgttgcccag gctggagtgc agtagcacga
1441 tctcggctca cttgcacctc cgtctctcgc ggttcaagca attgtctgcc tcagcctccc
1501 gagtagctgg gattacaggt gcgcactacc acgccaagct aatttttgta ttttttagta
1561 gagatggggt ttcaccatct tggccaggct ggtcttgaat tcctgacctc agtgatccac
1621 ccaccttggc ctcccaaaga tgctagtatt atgggcgtga accaccatgc ccagccgaaa
1681 agcttttgag gggctgactt caatccatgt aggaaagtaa aatggaagga aattgggtgc
1741 atttctagga cttttctaac atatgtctat aatatagtgt ttaggttctt tttttttca
1801 ggaatacatt tggaaattca aaacaattgg gcaaactttg tattaatgtg ttaagtgcag
1861 gagacattgg tattctgggc agcttcctaa tatgctttac aatctgcact ttaactgact
1921 taagtggcat taaacatttg agagctaact atatttttat aagactacta taaaaactac
1981 agagtttatg atttaaggta cttaaagctt ctatgggtga cattgtatat ataatttttt
2041 aaaaagggtt ttctatatgg ggattttcta tttatgtagg taatattgtt ctatttgtat
2101 atattgagat aatttattta atatacttta aataaagggt actgggaatt gtt (SEQ ID
NO: 111)

```

FIGURE 59A

103/115

TF (NM_001993)

METPAWPRVPRPETAVARTLLLGWVFAQVAGASGTTNTVAAYNL
TWKSTNFKTILEWEPKFVNQVYTVQISTKSGDWKSKCFYTTDTECDLTDEIVKDVKQT
YLARVFSYPAGNVESTGSAGEPLYENSPEFTPYLETNLGQPTIQSFEQVGTKVNVTV
DERTLVRRNNTFLSLRDVFGKDLIYTLYYWKSSSSGKKTAKTNTNEFLIDVDKGENYC
FSVQAVIPSRTVNRKSTDSPVECMGQEKGEFREIFYIIGAVVFVVIILVIILAI SLHK
CRKAGVGQSWKENSPLNVS (SEQ ID NO:112)

FIGURE 59B

104/115

GPR4 (NM_005282)

```

1  ctggtgacct tacttatctc tgttgctttc tggggtccta ggaaatgcc a gcactccac
61  ccacattgcc tgaactttcc aacactccct agctgcgctg tgcctatct caacacttcc
121 tcatgtatctt cttgtgtctt ctagaacatt ccccgccat tattacttca atatggctac
181 acatacttcc taattgccct gcaaaccatc tccttctcac cattgccag cgatgcttcc
241 gtctcctcca taaacactcc cggagaccaa tttttgtgtc accccatac tcctcgctg
301 acacactgac tccatacata acctccttga aaaacctctt tattaatctc accatcctcc
361 agacttccct cctgtcataa ttccatccct cctccaactt ttccctctca agctctgccc
421 ttcccagccc agcccagcct acccaacctc atctcttccc tgtagaccac atcccaccat
481 gttcccctga gcctccaagg aaggggctca gggggcccca tggcctcccg ctccctgtgg
541 ccccacagcc cccgtgggcc aggggaagcg ccccagaagc cgaagtgcc accatgggca
601 accacacgtg ggagggctgc cagctggact cgcgctgga ccacctctt ccgcatccc
661 tctacatctt tgtcatcggc gtggggctgc ccaccaactg cctggctctg tggggcgct
721 accgccaggt gcaacagcgc aacgagctgg gcgtctacct gatgaacctc agcatcgccg
781 acctgctgta catctgcacg ctgccgctgt ggggtgacta cttcctgcac cagcacaact
841 ggatccacgg ccccggttcc tgcaagctct ttgggttcat cttctacacc aatatctaca
901 tcagcatcgc cttcctgtgc tgcactctcg tggaccgcta cctggctgtg gccacccac
961 tccgcttcgc ccgctgcgc cgcgtcaaga ccgctgggc cgtgagctcc gtggctctgg
1021 ccacggagct gggcgccaac tcggcgcccc tgttccatga cgagctctc cgagaccgct
1081 acaaccacac cttctgcttt gagaagtcc ccatggaagg ctgggtggcc tggatgaacc
1141 tctatcggtt gttcgtgggc ttctcttcc cgtgggcgt catgctgctg tcgtaccggg
1201 gcatcctgcy ggcctgcyg ggcagcgtgt ccaccgagcy ccaggagaag gccaatatca
1261 agcggctggc cctcagcctc atcgccatcg tgctggctct ctttgcgccc tatcagctgc
1321 tcttgctgtc ccgcagcgcc atctacctgg gccgcccctg ggactgcggc ttcgaggagc
1381 ggcctcttcc tgcataccac agctcactgg ctttcaccag cctcaactgt gtggcgagcc
1441 ccactctcta ctgcctggtc aacgagggcg ccgcagcga tgtggccaag gccctgcaca
1501 acctgctccg ctttctggcc agcgacaagc ccaggagat ggccaatgcc tcgctcacc
1561 tggagacccc actcacctcc aagaggaaca gcacagccaa agccatgact ggcagctggg
1621 cggccactcc gccctcccag ggggaccagg tgcagctgaa gatgctgccc ccagcacaat
1681 gaaccccagag tggcacagaa tcccagttt tcccctctca tcccacagtc ccttctctcc
1741 tggctctggtg tatgcaaatt gtatggaaaa agggctgtgt taatattcat aagaatacaa
1801 gaacttagga agagttaggt tgggtgtgtc ctgggtcaacc tttgtgctcc cagatcccat
1861 cacagtttgg cgattgtgga gggcctcctg aaggaggaga tgagtaaata tatttttttg
1921 gagacagggc ctcactgtgt tgcccaggct ggagtgcagt agtgcagtgc tggctcactg
1981 cagcctccac ctccctgggc ctccagcgat cttcccatc cagcctcccg agtagctggg
2041 accacaaatg tgagcccacc catgcctggc taatttttgt actttttgt taaatggagt
2101 ctactatgt ttcccaggc tgatcttgaa ctccctgggc caagagatcc tcctgccttg
2161 gcctcccaaa gtgctcagat tagagatgtg agccgccatg tctggccaga taaattaagt
2221 caaacatttg gtttccagaa aataaagaca aatagagaag gttagatttt tttttttcca
2281 acaagtggat aaaagtctgt gactcggggg aaagtggaag gagaaatgca gccgatatag
2341 agtcattatg tttgcaaagc ccctggctat acaggccagg gaacataaga ccgcaattct
2401 aagtttctag ataaacagcg atctccaagt caagactgag gatgaagagg gagaatgtca
2461 gaactcaagt gaagggcaat cagggcagac tgcttgagg agtgatgcca gaaggttttg
2521 gaagaagggt tgggacaaga agaaagggt tttattcatt cattcaacag aggtttatgt
2581 agggcactgt gctgggtggg gctggggaca caacaatgac tgaggcagcc tggccttgcc
2641 ttcacagggc tcaccatata caagtaaata aaaaatatgt aatgtttgga attgct (SEQ

```

ID NO:113)

FIGURE 60A

105/115

GPR4 (NM_005282)

MGNHTWEGCHVDSRVDHLFPFSLYIFVIGVGLPTNCLALWAAAYR
QVQQRNELGVYLMNLSIADLLYICTLPLWVDYFLHHDNWIHGPGSCKLFGFIFYTNIY
ISIAFLCCISVDRYLAVAHPLRFARLRRVKTAVAVSSVWATELGANSAPLFHDELFR
DRYNHTFCFEKFFMEGWVAVWMNLYRVFVGFLFPWALMLLSYRGILRAVRGSVSTERQE
KAKIKRLALSLIAIVLVCFAPYHVLLLLSRSAIYLGRPWCDFEERVFSAYHSSLAFTS
LNCVADPILYCLVNEGARSVDKALHNLLRFLASDKPQEMANASLTLETPLTSKRNST
AKAMTGSWAATPPSQGDQVQLKMLPPAQ (SEQ ID NO:114)

FIGURE 60B

106/115

GPR66 (NM_006056)

```

1 agcgggggggt tcccggccgg acaggcgggg cgtcggggcg cgggctgggg ccgctgtcag
61 tcagtcactt ggtcccggc cgcgtctgt gtccgtcgt cggaggggtg aagccggggg
121 ctgcggggcc gcgggccgca tgactcctct ctgcctcaat tgctctgtcc tccttgga
181 cctgtaccca ggggggtgca ggaaccccat ggcttgcaat ggcagtgcgg ccagggggca
241 ctttgacctt gaggacttga acctgactga cgaggcactg agactcaagt acctggggcc
301 ccagcagaca gagctgttca tgcccatctg tgccacatac ctgctgatct tcgtggtggg
361 cgctgtgggg aatgggctga cctgtctggt catcctgcgc cacaaggcca tgcgcacgcc
421 taccaactac tacctcttca gcctggcctg gtccgacctg ctggtgctgc tgggtggcct
481 gcccttgag ctctatgaga tgtggcaca ctaccccttc ctgctgggcg ttggtggctg
541 ctatttccgc acgctactgt ttgagatggt ctgcctggcc tcagtgtcga acgtcactgc
601 cctgagcgtg gaacgctatg tggcgtggt gcacccactc caggccaggt ccattggtgac
661 gcggggccat gtgcgccgag tgcttggggc cgtctggggg cttgccatgc tctgtccct
721 gcccacacc agcctgcacg gcatccagca gctgcacgtg cctgcccggg gccagtgcc
781 agactcagct gtttgcatgc tggctccgcc acggggcctc tacaacatgg tagtgagac
841 caccgcgctg ctcttcttct gcctgcccat ggccatcatg agcgtgctct acctgctcat
901 tgggctgcga ctgcggcggg agaggctgct gctcatgcag gaggccaagg gcaggggctc
961 tgcagcagcc aggtccagat acacctgcag gctccagcag cactcaggg gccggagaca
1021 agtgaccaag atgctgtttg tcctggctgt ggtgtttggc atctgctggg cccggtcca
1081 cgccgaccgc gtcattgtga gcgtcgtgtc acagtggaca gatggcctgc acctggcctt
1141 ccagcacgtg cactcatctt ccggcatctt cttctacctg ggctcggcgg ccaaccccg
1201 gctctatagc ctcatgtcca gccgcttccg agagaccttc caggaggccc tgtgcctcgg
1261 ggctgctgc catcgccctc gaccccgcca cagctccac agcctcagca ggtatgacc
1321 aggcagacc ctgtgtgatg tgggctccct gggcagctgg tccatcctga gtggagcctt
1381 cgtggccca gaggcgcagc aagagaccga tccatcctga gtggagcctt aaagtggctt
1441 cacttgagg ggccagaggg tcacctggag ctggggagac acatctgcct tcctctgcag
1501 ggatccttca cgtactgtcc ctagttcagc ctagaaattc tgaccagcac ctcatgttcc
1561 ctgagaggga aacagcagga ggagggatcc ctgactgctg aggactcaca ctgaccagac
1621 gccacacctt gtgcttctta tctgtccact gccactcccc cagttcaaat ccttaccctg
1681 cagaaatata acagttagct ggggctcagc agtccctcct ctggggactc cctgccacca
1741 ctgccagttt ctgaaacggg cccactgggt cctcactgtc cttcccagtt cctgttcagg
1801 ttctggcagg ggcccaggga tccaggggac ctgggtccaa tctcagcctt gctgtcacca
1861 ccttgctcatg caccatcaag catatcagtc tacctttctt ttttctgag acagagtctc
1921 actctgtcgc ccaggctaga gtgcagtggc gcgatttttg ctcaactgaa cctccgcctc
1981 cgggggttcaa gcgattctcc tgctcagcc tcccagattg ctgggactac aggtgagccc
2041 cagcatgccc agctaatttt ttttaatttt tagtagagac ggggtttcac catgttggcc
2101 aggctggtct caaactcttg acctcaggtg atccgccgac ctccgctcc caaagtccctc
2161 ggattacagg catgagccac cacacccggc caatcagtc acctttctag gccttggttc
2221 cttgctgaa aatgaaaga ggcgtggct ttccacagtg tcatgctttg gcacttttagc
2281 tatggttttc tttctgtgtg tgtgtaagcc actgcttata ataaaacca caataccctc
2341 agactgaaag ggcggaagtt attatctgca tctttatcaa cccaagccc cacttctctc
2401 ctgacctccc catgcccctc ccagcctctc ccagcacaag tggggcacaag ccagcatgca
2461 agcagacccc accaccacag cccacctccg tcctcacata cgtgcaggct ggctcgggag
2521 tccagtgagc agagcattgg acttggtggt ccagaggggtc tctgagggca agagacatgg
2581 ccaaccaagg gcaaggagtg acctgtgga ggggtctgcc gaactcaatg cagtgagaag
2641 agggacaggg acaagtagtc cttgaaactg agccccattc tgaatccctg caggccaagt
2701 cattgctcag ccaggactca gttcatgggg gaaacttgac ctgctgcagt ccctgagctc
2761 tgtcctcctg agaggaagcc ctggcttcca aggctgggag ctggaggatg accttcgggtc
2821 ggtctgtctg ggttctccct gcagacagct tcctagctca tgcccatagc tcatgctccc
2881 tgccgagaaa gtggaggacg tggtagagg ttgcagatgt ttagttttaa aaattcaatt
2941 taaaaaataa taaatgctca tgatagaaaa tttggaaagt gcaaataagc aaaaatgaaa
3001 acaattttta aatgtaaaa cctctcttgc cagggaatgg ggggaaggga agtgaggagt
3061 tctttaatgg gtgaagagtt tcagttttgc aaaatgaaa agttctggag atcagttgtg
3121 caacaatatg aatatacata acaatactga actatacact gaaatgggta agatgggtaca
3181 ttttatgtta tgtgtatttt accacaattt ttataaaaag aggattaaat ctaaaggaaa
3241 gaaaaaatta aaaccaccca taactttact ctgaagcagt aacagtggca tgtttcctcc
3301 taaaaaaaaa aaaaaaaaaa gaagaaaaaa aaataaagaa aaaaaaaaaa aaaa (SEQ ID

```

NO:115)

FIGURE 61A

107/115

GPR66 (NM_006056)

MTPLCLNCSVLPGLYPGGARNPMA CNGSAARGHFDPEDLNLT
EALRLKYLGPQQTELFMPICATYLLIFVVGAVGNGLTCLVILRHKAMRTPTNYLFSL
AVSDLLVLLVGLPLELYEMWHNYPFLLGVGGCYFRTLLFEMVCLASVLNVTALSVERY
VAVVHPLQARSMVTRAHVRRVLGAVWGLAMLCSLPNTSLHGIQQLHVPCRGVPVPSAV
CMLVRPRALYNMVVQTTALLFFCLPMAIMSVLYLLIGLRLRRERLLLMQEAKGRGSAA
ARSRYTCRLQQHDRGRRQVTKMLFVLVVVFGICWAPFHADRVMSVVSQWTDGLHLAF
QHVHVISGIFFYLGSAANPVLYSLSMSSRFRETQFQALCLGACCHRLRPRHSSHSLSRM
TTGSTLCDVGSLGSWVHPLAGNDGPEAQQETDPS (SEQ ID NO:116)

FIGURE 61B

108/115

SLC22A2 (NM_003058)

```

1  ctttgaagtc agctggacca aggaaaggcc ctgccctgaa ggctggtcac ttgcagaggt
61 aaactcccc ctttgacttc tggccagggt ttgtgctgag ctggctgcag ccgctctcag
121 cctcgctccg ggcacgtcgg gcagcctcgg gccctcctgc ctgcaggatc atgccacca
181 ccgtggacga tgtcctggag catggagggg agtttcactt ttccagaag caaatgtttt
241 tcctcttggc tctgctctcg gctaccttcg cgcccatcta cgtgggcatc gtcttcctgg
301 gcttcacccc tgaccaccgc tgccggagcc ccggagtggc cgagctgagt ctgcgctgcg
361 gctggagtcc tgcagaggaa ctgaactaca cggtgccggg ccaggacct gcgggcgaag
421 cctccccaa acagtgtagg cgctacgagg tggactggaa ccagagcacc ttcgactgcg
481 tggaccccc gtccagcctg gacaccaaca ggagccgcct gccactgggc cctgcccggg
541 acggctgggt gtacgagacg cctggctcgt ccacgtcac cgagtttaac ctggtatgtg
601 ccaactcctg gatgttggac ctattccagt catcagttaa tgtaggatcc tttattggct
661 ctatgagtat cggctacata gcagacagggt ttggccgtaa gctctgcctc ctaactacag
721 tcctcataaa tgctgcagct ggagttctca tggccatttc cccaacctat acgtggatgt
781 taatttttcg cttaatccaa ggactggcca gcaaagcagg ctggttaata ggctacatcc
841 tgattacaga atttgttggg cggagatatc ggagaacagt ggggattttt taccaagttg
901 cctatacagt tgggctcctg gtgctagctg ggggtggctta cgcacttcct cactggaggt
961 gggtgcagtt cacagttgct ctgcccact tcttcttctt gctctattac tgggtcatac
1021 ctgagctctc caggtggctg atctcccaga ataagaatgc tgaagccatg agaatcatta
1081 agcacatcgc aaagaaaaat ggaaaatctc taccgcctc ccttcagcgc ctgagacttg
1141 aagaggaaac tggcaagaaa ttgaaccctt catttcttga cttggctcaga actcctcaga
1201 taaggaaaca tactatgata ttgatgtaca actgggtcac gagctctgtg ctctaccagg
1261 gcctcatcat gcacatgggc cttgcagggt acaatatcta cctggatttc ttctactctg
1321 ccctggttga attcccagct gcttccatga tcatcctcac catcgaccgc atcggacgcc
1381 gttacccttg ggctgcatca aatatggttg caggggcagc ctgtctggcc tcagttttta
1441 tacctggatga tctacaatgg ctaaaaatta ttatctcatg cttgggaaga atggggatca
1501 caatggccta tgagatagtc tgcttggtca atgctgagct gtacccaca ttcattagga
1561 atcttggcgt ccacatctgt tcctcaatgt gtgacattgg tggcatcatc acgccattcc
1621 tggctctacc gctcactaac atctggcttg agctcccgt gatggttttc ggcgtgcttg
1681 gcttggttgc tggaggtctg gtgctggtgc ttccagaaac taaagggaaa gctttgcctg
1741 agaccatcga ggaagccgaa aatatgcaaa gaccaagaaa aaataaagaa aagatgattt
1801 acctccaagt tcagaaacta gacattccat tgaactaaga agagagaccg ttgctgctgt
1861 catgacctag ctttgatggc agcaagacca aaagtagaaa tccctgcact catcacaag
1921 ccatacaac tcaaccaaac ttacccttga gccctatcaa cctaggtcta cagccagtgg
1981 agtctattgt aactgtgga aaaataccca tgggaccaga tcctgccaaa ttcttcagc
2041 tcactttatt ctcagcatte ctaggacatt ggacattggg tttctggagg gtttttttc
2101 catctttgta tttttttaa tttgattctt ttctttgcaa tgctatctaa ccagaataca
2161 taggggaact gtgggctagg caaacaatat agaaaaaagt gtgaaaaaca gtaagttgg
2221 gagaggagca tctattttct taaagaaata aaacacccaa aacaatataa agttgtccag
2281 aatgtatgtc aagaatttta gataggcctt tcagtaacac aggtgaagaa atttttaaaa
2341 atacattgat tattatctag gttagactta aagtgaatct caaataaaag aatcaggaat
2401 acaacttaag tgatcatgag gtccctccat atttagattg ggtaagcatg aatgtgtatt
2461 ttctacaaaa gaccttgaga agagttcaat aaaaaatgtt agcattataa aa (SEQ ID

```

NO:117)

FIGURE 62A

109/115

SLC22A2 (NM_003058)

MPTTVDDVLEHGGEFHFFQKQMFLLALLSATFAPIYVGIVFLG
FTPDHRCRSPGVAELSLRCGWSPAELNYTVPGPGPAGEASPRQCRRYEVDWNQSTFD
CVDPLASLDTNRSRLPLGPCRDGWVYETPGSSIVTEFNLVCANSWMLDLFQSSVNVGF
FIGSMSIGYIADRFGRKLCLLTTLINAAAGVLMASPTYTWMLIFRLIQGLVSKAGW
LIGYILITEFVGRRYRRRTVGIFYQVAYTVGLLVLAGVAYALPHWRWLQFTVALPNFFF
LLYYWCIPESPRWLISQNKNAEAMRIIKHIAKKNKSLPASLQRLRLEEETGKKLNPS
FLDLVRTPQIRKHTMILMYNWF TSSVLYQGLIMHMGLAGDNIYLDFFYSALVEFPAAF
MIILTIDRIGRRYPWAASNMVAGAACLASVFIPGDLQWLKIIISCLGRMGITMAYEIV
CLVNAELYPTFIRNLGVHICSSMCDIGGIITPFLVYRLTNIWLELPLMVFGVLGLVAG
GLVLLL PETKGKALPETIEEAENMQRPKNKEKMIYLVQVQKLDIPLN (SEQ ID NO:118)

FIGURE 62B

110/115

NLSN1 (NM_002420)

```

1 gccctggcca aggaggaggc tgaaagagcc tgagctgtgc cctctccatt ccactgctgt
61 ggcagggtca gaaatcttgg atagagaaaa ccttttgcaa acgggaatgt atctttgtaa
121 ttcctagcac gaaagactct aacagggtgtt gctgtggcca gttcaccaac cagcatatcc
181 cccctctgcc aagtgcacaa cccagcaaaa atgaagagga aaacaaacag gtggagactc
241 agcctgagaa atggtctgtt gccaagcaca cccagagcta cccaacagat tcctatggag
301 ttcttgaatt ccagggtggc ggatattcca ataaagccat gtatatccgt gtatcctatg
361 acaccaagcc agactcactg ctccatctca tggtgaaaga ttggcagctg gaactcccca
421 agctcttaat atctgtgcat ggaggcctcc agaactttga gatgcagccc aagctgaaac
481 aagtccttgg gaaaggcctg atcaaggctg ctatgaccac cggggcctgg atcttcaccg
541 ggggtgtcag cacagggtgtt atcagccacg taggggatgc cttgaaagac cactcctcca
601 agtccagagg ccgggtttgt gctataggaa ttgctccatg gggcatcgtg gagaataagg
661 aagacctggg tggaaaggat gtaacaagag tgtaccagac catgtccaac cctctaagta
721 agctctctgt gctcaacaac tcccacccc acttcaccc ggctgacaat ggcaccctgg
781 gcaagtatgg cgccgaggtg aagctgcgaa ggctgctgga aaagcacatc tccctccaga
841 agatcaacac aagactgggg cagggcgtgc ccctcgtggg tctcgtggg gaggggggcc
901 ctaacgtggg gtccatcgtc ttggaatacc tgcaagaaga gcctcccatc cctgtggtga
961 tttgtgatgg cagcggacgt gcctcggaca tcctgtcctt tgccgacaag tactgtgaag
1021 aaggcggaa aataaatgag tccctcaggg agcagcttct agttaccatt cagaaaacat
1081 ttaattataa taaggcacaa tcacatcagc tgtttgcaat tataatggag tgcataga
1141 agaaagaact cgtcactgtg ttcagaatgg gttctgaggg ccagcaggac atcgagatgg
1201 caattttaac tgccctgctg aaaggaacaa acgtatctgc tccagatcag ctgagcttgg
1261 cactggcttg gaaccgctg gacatagcac gaagccagat ctttgtctt gggccccact
1321 ggccgcccc tgggaagcctg gcacccccga cggacagcaa agccacggag aaggagaaga
1381 agccacccat ggccaccacc aaggaggagg gaggaaaagg gaaaggcaag aagaaagggg
1441 aagtgaagaa ggaagtggag gaagaaactg acccccggaa gatagagctg ctgaactggg
1501 tgaatgcttt ggagcaagcg atgctagatg ctttagtctt agatcgtgtc gactttgtga
1561 agtcctgat tgaaaacgga gtgaacatgc aacactttct gaccattccg aggctggagg
1621 agctttataa cacaagactg ggtccaccaaa acacacttca tctgctggg agggatgtga
1681 aaaagagcaa ccttcgcctt gattaccaca tcagcctcat agacatcggg ctgctgctgg
1741 agtacctcat gggaggagcc taccgctgca actacactcg gaaaaacttt cggacccttt
1801 acaacaactt gtttggaacca aagaggccta aagctcttaa acttctggga atggaagatg
1861 atgagcctcc agctaaaggg aagaaaaaaa aaaaaaagaa aaaggaggaa gagatcgaca
1921 ttgatgtgga cgaccctgcc gtgagtcggt tccagtatcc cttccacgag ctgatgggtg
1981 gggcagtgct gatgaaacgc cagaaaatgg cagtgttctt ctggcagcga ggggaagaga
2041 gcatggccaa ggccctgggtg gcctgcaagc tctacaaggc catggcccac gactcctccg
2101 agagtgatct ggtggatgac atctcccagg acttgataaa caattccaaa gacttcggcc
2161 agcttgcttt ggagttatta gaccagtctt ataagcatga cgagcagatc gctatgaaac
2221 tcctgaccta cgagctgaaa aactggagca actcgacctg cctcaaaact gccgtggcag
2281 ccaaacaccg ggacttcatt gctcacacct gcagccagat gctgtgacc gatagtggg
2341 tgggaagact gcggatgcgg aagaaccccc gctgaagggt tatcatgggg attcttctac
2401 ccccaccat cttgtttttg gaatttcgca catatgatga tttctcgtat caaacatcca
2461 aggaaaacga ggatggcaaa gaaaaagaag aggaaaatac ggatgcaaat gcagatgctg
2521 gctcaagaaa gggggatgag gagaacgagc ataaaaaaca gagaagtatt cccatcggaa
2581 caaagatctg tgaattctat aacgcgcoca ttgtcaagtt ctgggttttac acaatatcat
2641 acttgggcta cctgctgctg ttttaactacg tcatcctggt gcggatggat ggctggccgt
2701 ccctccagga gtggatcgtc atctcctaca tcgtgagcct ggcgttagag aagatacgag
2761 agatcctcat gtcagaacca ggcaaactca gccagaaaat caaagtttgg cttcaggagt
2821 actggaacat cacagatctc gtggccattt ccacattcat gattggagca gattctcgcc
2881 tacagaacca gccctacatg ggctatggcc gggtgatcta ctgtgtggat atcatcttct
2941 ggtacatccg tgcctggac atcttttggt tcaacaagta tctggggcca tacgtgatga
3001 tgattggaaa gatgatgatc gacatgctgt actttgtggg catcatgctg gtcgtgctca
3061 tgagtttcgg agtagcccg caagccattc tgcattcaga ggagaagccc tcttggaac
3121 tggcccgaaa catcttctac atgccctact ggatgatcta tggagaggtg tttgcagacc
3181 agatagacct ctacgccatg gaaattaatc ctcttgtgg tgagaaccta tatgatgagg
3241 agggcaagcg gcttcctccc tgtatccccg gcgcctggct cactccagca ctcatggcgt
3301 gctatctact ggtcgccaac atcctgctgg tgaacctgct gattgctgtg ttcaacaata

```

FIGURE 63A

111/115

```

3361 ccttcttttga agtaaaatca atatccaacc aggtgtggaa gttccagcga tatcagctga
3421 ttatgacatt tcatgacagg ccagtcctgc cccacccgat gatcatttta agccacatct
3481 acatcatcat tatgcgtctc agcggccgct gcaggaaaaa gagagaaggg gaccaagagg
3541 aacgggatcg tggattgaag ctcttcctta gcgacgagga gctaaagagg ctgcatgagt
3601 tcgaggagca gtgcgtgcag gagcacttcc gggagaagga ggatgagcag cagtcgtcca
3661 gcgacgagcg catccgggtc acttctgaaa gagttgaaaa tatgtcaatg aggttggaag
3721 aaatcaatga aagagaaact tttatgaaaa cttccctgca gactgttgac cttcgacttg
3781 ctcagctaga agaattatct aacagaatgg tgaatgctct tgaaaatctt gcgggaatcg
3841 acaggtctga cctgatccag gcacgggtccc gggcttcttc tgaatgtgag gcaacgtatc
3901 ttctccggca aagcagcatc aatagcgctg atggctacag cttgtatcga tatcatttta
3961 acggagaaga gttattattt gaggatacat ctctctccac gtcaccaggg acaggagtca
4021 ggaaaaaaac ctgttccttc cgtataaagg aagagaagga cgtgaaaacg cacctagtcc
4081 cagaatgtca gaacagtctt cacccttcac tgggcacaag cacatcagca accccagatg
4141 gcagtcacct tgcagtagat gacttaaaga acgctgaaga gtcaaaatta ggtccagata
4201 ttgggatttc aaaggaagat gatgaaagac agacagactc taaaaaagaa gaaactattt
4261 ccccaagttt aaataaaaca gatgtgatac atggacagga caaatcagat gttcaaaaca
4321 ctcagctaac agtggaaacg acaaatatag aaggcactat ttcctatccc ctggaagaaa
4381 ccaaaaattac acgctatttc cccgatgaaa cgatcaatgc ttgtaaaaca atgaagtcca
4441 gaagcttcgt ctattcccgg ggaagaaagc tggtcggtgg ggttaaccag gatgtagagt
4501 acagttcaat cacggaccag caattgacga cggaatggca atgccaagtt caaaagatca
4561 cgcgctctca tagcacagat attccttaca ttgtgtcgga agctgcagtg caagctgagc
4621 ataaagagca gtttgagat atgcaagatg aacaccatgt cgctgaagca attcctcgaa
4681 tccctcgctt gtccctaacc attactgaca gaaatgggat ggaaaactta ctgtctgtga
4741 agccagatca aactttggga ttcccatctc tcaggtcaaa aagtttacat ggacatccta
4801 ggaatgtgaa atccattcag ggaaagttag acagatctgg acatgccagt agtgtaagca
4861 gcttagtaat tgtgtctgga atgacagcag aagaaaaaaa ggttaagaaa gagaaagctt
4921 ccacagaaac tgaatgctag tctgttttgt ttctttaatt ttttttttta acagtcagaa
4981 ccaactaatgg gtgtcatctt ggccatctaa acatcatcaa tttctaaaaa cattttccct
5041 taaaaaattt tggaaattca gacttgattt acaatttaat gcactaaaag tagtattttg
5101 ttagcatatg ttagtaggct tagttttttc agttgcagta gtatcaaatg aaagtgatga
5161 tactgtaacg aagataaatt ggctaatacag tatacaagat tatacaatct ctttattact
5221 gagggccacc aaatagccta ggaagtgcc tcgagcactg aagtcaccat taggtcactt
5281 aagaagtaag caactagctg ggcacagtgg ctcatgcctg taatcctagc actttgggag
5341 gccaaaggcag aaagatagct tgagtccagg agtttgagac cagcctgggc aacatagtga
5401 taccatct cttaaaaaaa aaaaaaaaaa a (SEQ ID NO:119)

```

FIGURE 63B

112/115

NLSN1 (NM_002420)

MYIRVSYDTKPDSSLHLMVKDWQLELPKLLISVHGGLQNFEMQP
KLKQVFGKGLIKAAMTTGAWIFTGGVSTGVISHVGDALKDHSSKSRGRVCAIGIAPWG
IVENKEDLVGKDVTRVYQTMNSPLSKLSVLNNSHTHFILADNGTLGKYGAEVKLRRLI
EKHISLQKINTRLGQGVPLVGLVVEGGPNVVSIVLEYLQEEPPIPVVICDGSGRASDI
LSFAHKYCEEGLIINESLREQLLVTIQKTFNYNKAQSHQLFAIMECKKKELVTVFR
MGSEQQODIEMAILTALLKGTNVSAPDQLSLALAWNVRVDIARSQIFVFGPHWPPLGSL
APPTDSKATEKEKKPPMATTKGGRGKGGKGGKGVKEEVEEETDPRKIELLNWVNALE
QAMLDALVLDVRDVFVKLLIENGVMQHFLLTIPLREELYNTRLGPPNTLHLLVRDVKKS
NLPPDYHISLIDIGLVLEYLMGGAYRCNYTRKNFRTLNNLFGPKRPPKALKLLGMEDD
EPPAKGKKKKKKKEEIDIDVDDPAVSRFQYPFHELMVWAVLMKRQKMAVFLWQRGE
ESMAKALVACKLYKAMAHESSES DLVDDISQDLNNSKDFGQLALELLDQSYKHDEQI
AMKLLTYELKNWSNSTCLKLVAACHRDFIAHTCSQMLLTDMMGRRLMRKNPGLKVI
MGILLPPTILFLEFRTYDDFSYQTSKENEDGKEKEEENTDANADAGSRKGDEENEHKK
QRSIPIGTKICEFYNAPIVKFWFYTISYLGYLFFFNYVILVRMDGWPSLQEWIVISYI
VSLALEKIREILMSEPGKLSQKIKVWLQEYWNITDLVAISTFMIGAILRLQNQPYMGY
GRVIYCVDIIFWYIRVLDIFGVNKYLGYPVMMIGKMMIDMLYFVIMLVVLMVSGVAR
QAILHPEEKPSWKLARNIFYMPYWMIYGEVFADQIDLYAMEINPPCGENLYDEEGKRL
PPCIPGAWLTPALMACYLLVANILLVNLLIAVFNNNTFFEVKSSISNQVWKFORYLIMT
FHDRPVLPPPMIILSHIYIIIMRLSGRCRKKREGDQEERDRGLKLFLSDEELKRLHEF
EEQCQVEHFREKEDEQQSSDERIRVTSEVENMSMRLEEINERETFMKTSLQTVDLR
LAQLEELSNRMVNALENLAGIDRSDLIQARSASSECEATYLLRQSSINSADGYSLYR
YHFNGEELLFEDTSLSTSPGTGVRKKTCSFRIKEEKDVKTHLVPECQNSLHLSLGTST
SATPDGSHLAVDDLKNAEESKLGPDIGISKEDDERQTD SKKEETISPSLNKTDVIHQ
DKSDVQNTQLTVETTNIEGTISYPLEETKITRYFPDETINACKTMKSRSFVYSRGRKL
VGGVNQDVEYSSITDQQLTTEWQCQVQKITRSHSTDIPIYVSEAAVQAEHKEQFADMQ
DEHHVAEAIPIRSLTITDRNGMENLLSVKPDQTLGFPSLRKSLHGHPNVKSIQ
GKLDRSGHASSVSSLVIVSGMTAEKKVKKEKASTETEC (SEQ ID NO:120)

FIGURE 63C

113/115

ATN2 (Na/K transport, NM_000702)

```

1  tctctgtctg ccaggggtctc cgactgtccc agacgggctg gtgtgggctt gggatcctcc
61  tgggtgacctc tcccgcctaag gtccctcagc cactctgccc caagatgggc cgtggggctg
121  gccgtgagta ctcacctgcc gccaccacgg cagagaatgg gggcggaag aagaaacaga
181  aggagaagga actggatgag ctgaagaagg aggtggcaat ggatgaccac aagctgtcct
241  tggatgagct gggccgcaaa taccaagtgg acctgtccaa gggcctcacc aaccagcggg
301  ctcaggacgt tctggctcga gatgggcccc acgccctcac accacctccc acaaccctg
361  agtgggtcaa gttctgccgt cagcttttcg ggggggttctc catcctgctg tggattgggg
421  ctatcctctg cttcctggcc tacggcatcc aggtctgcat ggaggatgaa ccatccaacg
481  acaatctata tctgggtgtg gtgctggcag ctgtggtcat tgtcactggc tgccttctct
541  actaccagga ggccaagagc tccaagatca tggattcctt caagaacatg gtacctcagc
601  aagcccttgt gatccgggag ggagagaaga tgcagatcaa cgcagaggaa gtgggtgggg
661  gagacctggt ggaggtgaag ggtggagacc gcgtccctgc tgacctccgg atcatctctt
721  ctcatggctg taaggtggat aactcatcct taacaggaga gtcggagccc cagaccctgt
781  ccccgagtt caccatgag aacccctgg agaccgcaa tatctgtttc ttctccacca
841  actgtgttga aggcactgcc aggggcattg tgattgccac aggagaccgg acggtgatgg
901  gccgcatagc tactctgcc tcaggcctgg aggttggggc gacaccata gcaatggaga
961  ttgaacactt catccagctg atcacagggg tcgctgtatt cctgggggtc tcttcttctg
1021  tgctctccct catcctgggc tacagctggc tggaggcagt catcttctct atcggcatca
1081  tagtggccaa cgtgcctgag gggcttctgg ccactgtcac tgtgtgcctg accctgacag
1141  ccaagcgcac ggcacggaag aactgcctgg tgaagaacct ggaggcgggt gagacgctgg
1201  gctccacgtc caccatctgc tcggacaaga cgggcaccct caccagaac cgcatgaccg
1261  tcgccacat gtggttcgac aaccaaactc atgaggctga caccacgaa gatcagtctg
1321  gggccacttt tgacaaacga tccctacgt ggacggccct gtctcgaatt gctggtctct
1381  gcaaccgcgc cgtcttcaag gcaggacagg agaacatctc cgtgtctaag cgggacacag
1441  ctggtgatgc ctctgagtca gctctgctca agtgcattga gctctcctgt ggctcagtga
1501  ggaaaatgag agacagaaac cccaagggtg cagagattcc tttcaactct accaacaagt
1561  accagctgtc tatccacgag cgagaagaca gccccagag ccacgtgctg gtgatgaagg
1621  gggccccaga gcgcattctg gaccgggtgt gatgcctttc aaaatgccta catggagctg gggggacttg
1681  cgctcgacaa ggagatgcaa gatgcctttc atctgccatc tggaaagttt cctcggggct
1741  gggagcgtgt gctgggattc tgtcaactga atctgccatc tggaaagttt cctcggggct
1801  tcaaattcga cacggatgag ctgaactttc ccacggagaa gctttgcttt gtggggctca
1861  tgtctatgat tgacctccc cgggtgctg tgccagatgc tgtgggcaag tgccgaagcg
1921  caggcatcaa ggtgatcatg gtaaccgggg atcacctat cacagccaag gccattgcca
1981  aaggcgtggg catcatatca gagggtaacg agactgtgga ggacattgca gcccggtca
2041  acattcccat gagtcaagtc aacccagag aagccaaggc atgctggtg cacggctctg
2101  acctgaagga catgacatcg gagcagctcg atgagatcct caagaaccac acagagatcg
2161  tctttgctcg aacgtctccc cagcagaagc tcatcattgt ggagggatgt cagaggcagg
2221  gagccattgt ggccgtgacg ggtgacgggg tgaacgactc cctgcatatg aagaaggctg
2281  acattggcat tgccatgggc atctctggct ctgacgtctc taagcaggga gccgacatga
2341  tcctgctgga tgacaacttt gctccatcg tcacgggggt ggagggggc ggcctgatct
2401  ttgacaactt gaagaaatcc atcgccatca ccctgaccag caacatcccc gagatcacc
2461  ccttctgct gttcatcatt gccaacatcc ccctacctct gggcactgtg accatccttt
2521  gcattgacct gggcacagat atggtccctg ccatctcctt ggcctatgag gcagctgaga
2581  gtgatcatc gaagcggcag ccacgaaact ccagacgga caagctggtg aatgagaggc
2641  tcatcagcat ggcctacgga cagatcggga tgatccaggc actgggtggc ttcttcacct
2701  actttgtgat cctggcagag aacggtttcc tgccatcacg gctactggga atccgcctcg
2761  actgggatga ccggaccatg aatgatctgg aggacagcta tggacaggag tggacctatg
2821  agcagcgga ggtggtggag ttacgtgcc acacggcatt ctttgccagc atcgtggtg
2881  tgagtgggc tgacctcgc atctgcaaga ccgcgcgcaa ctcagttctc cagcagggca
2941  tgaagaacaa gatcctgatt tttgggtctc tggaggagac ggcgttggct gccttctct
3001  ctactgcc cccctacagc ctctcatct tcatctatga tgaggctcga aagctcatcc
3061  tctgcgcctt cctggtggc tgggtggaga aggagacata ctactgacc cattggaaga
3121  tgccggcgta tcctggtggc tgggtggaga ggggagctct ggaggtgttg tggggatgg
3181  agaaccaggc atggaaagat ggggagctct ggaggtgttg agagataatg agggcaactc
3241  gatggaaata acgggtggca ttgggtggca acatttgggg

```

FIGURE 64A

114/115

```

3301 gcaggctaag ttgcggggta tataaattgg ggtgatgacc ccatagacct aactgtgaac
3361 aatcagatta gacactatgt gttagagtcc ccccgaccag atccttttcc atcccactcc
3421 actatgttgt ctattttttc tgaggaatta agggttaccc caccctgccc actcccactcc
3481 cttcaacccc acttcctact gtaatagatc agcatccaaa agcaggaacc catctaaacc
3541 agaaggaagc cctctcagat caccocagcc tcaactccatt tcccacttcc acccccgtta
3601 gcttcctgca ggactctatc cctggcttcc ccttcagacc ttgcaatcac aaaagggttct
3661 tctgggtgagt gcaagagcct gagactggaa aagggtggact tgtctcccag tccaggctgg
3721 taagggacct tcagggagag ctgggcagac aggtgggaga tggaggtagg gctggctgga
3781 ggaaggaaac aacaaaggaa gtgaggtagt gccaatgaca ggacatttga catgagtctc
3841 cagatagatg tcgtggactc cagctctacg tcccacattt tagaatacc caccagcaga
3901 acaaaactcag atctcatcag ggtagcagca gaggcaggac cagaaggcaa tcaagagctt
3961 ccagaaatgc cacacttgtg tgccacagag ttccccgctg acccttggtt aggggtcctc
4021 ttagtccaca aggtccggat gtcactcatg tacttaataa cacttcacct tctgtaatac
4081 taagtccctc gagctccatg ctgttctgaa agggatggcc acaagttctt tcccagcctc
4141 ttccattccc tttcttttca tgcccatccc gatgaacctg catcattccc cgacactgcc
4201 aagccaaccc tggaaaagga gttcgctggc cattggctag aatcaggggtg gagaagttcc
4261 ctgaaccttc ctgtctccca gggacatgta tgcttccagg gacaagctta ggtcatgaac
4321 atggtcagaa cctttggaca agaggaaaaa tactaagaga tttgcttttt ctgggtgcgg
4381 tggctcatgc ctgtaatccc agcactttgg gaggccgagg caggtggatc atgaggtcag
4441 gagttcgagg cgagcctggc caacatggtg aaacctgtc tctactaaaa gtacaaaaaa
4501 ttagccagtc atgggtggc acgctgtaa tctcagctac tcaggaggct gaggcaggag
4561 aattgcttga acctgtgagg aagaggttgc agtgagctga gatcgtgcc ttacactcca
4621 gcctgggcga aagggtgaga ctccatctca aaaaaaaaaa aaatgatttg cttttgacgt
4681 cttaggtggc agggctgttc cctccaggca aatgcccttc aaaccgacga tcattgtgcc
4741 cacttaccct gggctggaga gttggtttca ggcttctaca ggagatagct ttctttccct
4801 tactccctat ctaacacttt tgctctgcag gcagccttgc ccattctcta agcctggctt
4861 agaaggcact gggaatgtcc tgtagagaga gacctagata ggtcatgcaa gtgagaaaga
4921 catctgagga aaatggaaga cctaaggcag acaggaagga agcacaaaag acaagcattg
4981 ggtcagaccc ataaaccacc tcccaaaggc tgtcatttca ttgactgga attttgcttt
5041 atcagaagca aggaagtaag ggagtcattg ccttgggcct gggaatctaa gtgggagaca
5101 atattaattt ggatccgatt aattggagat tactaactgt ggacaaaagt ttatctttgc
5161 acaatcaata aaaatggcat ttttttagta aattaagagc ataaacaata ttgctagagg
5221 tggcatgttt agtctaccaa aaacaatact tttcaggcac tttagaaata tctttttaga
5281 agcagcgagt gcatgggcta attatcatca atctttatgt atttgttaaa gaaacatcta
5341 caggatcttt attgggtgacc ttttgtaaga cattagtttg aggtactacc tatctacttg
5401 aaaataataa agtggcattt ctttatgaaa aaaaaagaaa tctcttccat aattcagatt
5461 tctacacttt atacttgcc cctcctaaa tcgtgatatt gaaatatggt g (SEQ ID

```

NO: 121)

FIGURE 64B

115/115

ATN2 (Na/K transport, NM_000702)

MGRGAGREYSPAATTAENGGGKKKQKEKELDELKKEVAMDDHKL
SLDELGRKYQVDLSKGLTNQRAQDVLARDGPNALTPPPTTPEWVKFCRQLFGGFSILL
WIGAILCFLAYGIQAAMEDEPSNDNLYLGVVLAADVIVTGCFSYQEAKSSKIMDSFK
NMVPQQALVIREGEKMQINAEVVGDLVEVKGGDRVPADLRIISSHGCKVDNSSLTG
ESEPQTRSEFTHENPLETRNICFFSTNCVEGTARGIVIATGDRTVMGRIATLASGLE
VGRTPIAMIEIHFILITGVAVFLGVSFVLSLILGYSWLEAVIFLIGIIVANVPEGL
LATVTVCLTLTAKRMARKNCLVKNLEAVETLGSTSTICSDKTGTLTQNRMTVAHMFWD
NQIHEADTTEDQSGATFDKRSPTWTALSRIAGLCNRAVFKAGQENISVSKRDTAGDAS
ESALLKCIELSCGSVRKMRDRNPKVAEIPFNSTNKYQLSIHEREDSPQSHVLVMKGAP
ERILDRCSLILVQGKEIPLDKEMQDAFQONAYMELGGLGERVLGFCQLNLP SGKFPRGF
KFDTDELNFPTEKLCFVGLMSMIDPPRAAVPDAVGKCRSAGIKVIMVTGDHPITAKAI
AKGVGIISEGNETVEDIAARLNIPMSQVNPREAKACVVHGSDLKDMTSEQLDEILKNH
TEIVFARTSPQOKLIIIVEGCQRQGAIVAVTGDGVNDSPALKKADIGIAMGISGSDVSK
QAADMILLDDNFASIVTGVEEGRLI FDNLKKSIAYTLTSNIPEITPFLLFIIANIPLP
LGTVTILCIDLGTDMVPAISLAYEAAESDIMKRQPRNSQTDKLVNERLISMAYGQIGM
IQALGGFFTYFVILAENGFLPSRLLGIRLDWDDRTMNDLEDSYGQEWTYEQRKVVEFT
CHTAFFASIVVQWADLIICKTRRNSVFQQGMKNKILIFGLLEETALAFLSYCPGMG
VALRMYPLKVTTWWFCAFPYSLLIIFIYDEVRLILRRYPGGWVEKETYY (SEQ ID NO:122)

FIGURE 64C

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
27 May 2004 (27.05.2004)

PCT

(10) International Publication Number
WO 2004/044178 A3

(51) International Patent Classification⁷: **C07H 21/04**

(21) International Application Number:
PCT/US2003/036260

(22) International Filing Date:
13 November 2003 (13.11.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/425,813 13 November 2002 (13.11.2002) US

(71) Applicant (*for all designated States except US*): **GENENTECH, INC.** [US/US]; 1 DNA Way, South San Francisco, CA 94080-4990 (US).

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): **SMITH, Victoria** [AU/US]; 19 Dwight Road, Burlingame, CA 94010 (US).

(74) Agents: **CONLEY, Deirdre L.** et al.; GENENTECH, INC., 1 DNA Way, South San Francisco, CA 94080-4990 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(88) Date of publication of the international search report:
22 September 2005

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND COMPOSITIONS FOR DIAGNOSING DYSPLASIA

(57) Abstract: Methods and compositions are disclosed for detecting dysplasia in a tissue sample, screening candidate compounds for the ability to inhibit growth of a cancer cell, predicting predisposition to adenocarcinoma and treating cancer based on gene expression profiles.



WO 2004/044178 A3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US03/36260

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/04

US CL : 536/23.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

STN: EMBASE BIOSIS CAPLUS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 0206526 (UNIV CALIFORNIA) 24 January 2002.	1-30, 37-45

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"

document member of the same patent family

Date of the actual completion of the international search

25 June 2005 (25.06.2005)

Date of mailing of the international search report

14 JUL 2005

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US
Commissioner of Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Facsimile No. (703) 305-3230

Authorized officer

Celine Qian, Ph.D.

Telephone No. 571-273-8300

CELIN QIAN
PATENT EXAMINER

Hella Celine Qian

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US03/36260

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claim Nos.: 31-36
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
The claims cannot be searched because the CRF is defect.
3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐
☐

- The additional search fees were accompanied by the applicant's protest.
No protest accompanied the payment of additional search fees.